

Novel Characteristics of Split Trees by use of Renewal Theory

Cecilia Holmgren

Abstract

We investigate characteristics of random split trees introduced by Devroye [3]; split trees include for example binary search trees, m -ary search trees, quadrees, median of $(2k + 1)$ -trees, simplex trees, tries and digital search trees. More precisely: We introduce the use of renewal theory in the studies of split trees, and use this theory to prove several results about split trees. A split tree of cardinality n is constructed by distributing n “balls” (which often represent “key numbers”) in a subset of vertices of an infinite tree. One of our main results is to give a relation between the deterministic number of balls n and the random number of vertices N . In [3] there is a central limit law for the depth of the last inserted ball so that most vertices are close to $\frac{\ln n}{\mu} + \mathcal{O}(\sqrt{\ln n})$, where μ is some constant depending on the type of split tree; we sharpen this result by finding an upper bound for the expected number of vertices with depths $\geq \frac{\ln n}{\mu} + \ln^{0.5+\epsilon} n$ or depths $\leq \frac{\ln n}{\mu} + \ln^{0.5+\epsilon} n$ for any choice of $\epsilon > 0$. We also find the first asymptotic of the variances of the depths of the balls in the tree.

1 Introduction

1.1 Preliminaries

In this paper we consider random split trees introduced by Devroye [3]. Some important examples of split trees are binary search trees, m -ary search trees, quadrees, median of $(2k + 1)$ -trees, simplex trees, tries and digital search trees. As shown in [3] the split trees belong to the family of so-called $\log n$ trees, i.e., trees with height (maximal depth) *a.a.s.* $\mathcal{O}(\log n)$. (For the notation *a.a.s.*, see [13].)

The (random) split trees constitute a large class of random trees which are recursively generated. Their formal definition is given in the “split tree

generating algorithm” below. To facilitate the penetration of this rather complex algorithm we will first provide a brief heuristic description.

A skeleton tree S_b of branch factor b is an infinite rooted tree in which each vertex has exactly b children that are numbered $1, 2, \dots, b$. A split tree is a finite subtree of a skeleton tree S_b . The split tree is constructed recursively by distributing balls one at a time to a subset of vertices of S_b . We say that the tree has cardinality n if n balls are distributed. Since many of the common split trees come from algorithms in Computer Science the balls often represent some “key numbers” or other data symbols. There is also a so-called vertex capacity, $s > 0$, which means that each node can hold at most s balls. We say that a vertex v is a leaf in a split tree if the node itself holds at least one ball but no descendants of v hold any balls. The split tree consists of the leaves and all the ancestors of the leaves, in particular the root of S_b , but no descendant of a leaf is included. In this way the definition of leaves in split trees is equivalent to the usual definition of leaves in trees. See Figure 1 and Figure 2, where two examples of split trees are illustrated (the parameters s_0 and s_1 in the figures are introduced in the formal “split tree generating algorithm”).

The first ball is placed in the root of S_b . A new ball is added to the tree by starting at the root, and then letting the ball fall down to lower levels in the tree until it reaches a leaf. Each vertex v of S_b is given an independent copy of the so-called random split vector $\mathcal{V} = (V_1, V_2, \dots, V_b)$ of probabilities, where $\sum_i V_i = 1$ and $V_i \geq 0$. The split vectors control the path that the ball takes until it finally reaches a leaf; when the ball falls down one level from vertex v to one of its children, it chooses the i -th child of v with probability V_i , i.e., the i -th component of the split vector associated to v . When a full leaf (i.e., a leaf which already holds s balls) is reached by a new ball it splits. This means that some of the $s + 1$ balls are given to its children, leading to new leaves so that more nodes will be included in the tree. When all the n balls are distributed we end up with a split tree with a finite number of nodes which we denote by the parameter N .

The split tree generating algorithm: The formal, comprehensive “split tree generating algorithm” is as follows with the following introductory notation. The (random) split tree has the parameters b, n, s and \mathcal{V} as we described above; there are also two other parameters: s_0, s_1 (related to the parameter s) that occur in the algorithm below. Let n_v denote the total number of balls that the vertices in the subtree rooted at vertex v hold together, and C_v be the number of balls that are held by v itself. Thus, we note that a vertex v is a leaf if and only if $C_v = n_v > 0$. Also note that a vertex $v \in S_b$ is included in the split tree if, and only if, $n_v > 0$. If $n_v = 0$, the vertex v is not included and it is called useless.

Below there is a description of the algorithm which determines how the n balls are distributed over the vertices. Initially there are no balls, i.e., $C_v = 0$ for each vertex v . Choose an independent copy \mathcal{V}_v of \mathcal{V} for every vertex $v \in S_b$. Add balls one by one to the root by the following recursive procedure for adding a ball to the subtree rooted at v .

1. If v is not a leaf, choose child i with probability V_i , and recursively add the ball to the subtree rooted at child i , by the rules given in steps 1, 2 and 3.
2. If v is a leaf and $C_v = n_v < s$, (s is the capacity of the vertex) then add the ball to v and stop. Thus, C_v and n_v increase by 1.
3. If v is a leaf and $C_v = n_v = s$, the ball cannot be placed at v since it is occupied by the maximal number of balls it can hold. In this case let $n_v = s + 1$ and $C_v = s_0$, by placing $s_0 \leq s$ randomly chosen balls at v and $s + 1 - s_0$ balls at its children. This is done by first giving s_1 randomly chosen balls to each of the b children. The remaining $s + 1 - s_0 - bs_1$ balls are placed by choosing a child for each ball independently according to the probability vector $\mathcal{V}_v = (V_1, V_2, \dots, V_b)$, and then using the algorithm described in steps 1, 2 and 3 applied to the subtree rooted at the selected child. Note that if $s_0 > 0$ or $s_1 > 0$, this procedure does not need to be repeated since no child could reach the capacity s , whereas in the case $s_0 = 0$ this procedure may have to be repeated several times.

From 3. it follows that the integers s_0 and s_1 have to satisfy the inequality

$$0 \leq s_0 \leq s, \quad 0 \leq bs_1 \leq s + 1 - s_0.$$

Note that every nonleaf vertex has $C_v = s_0$ balls and every leaf has $0 < C_v \leq s$ balls.

Figure 1 shows a split tree with cardinality 32 and parameters $(b, s, s_0, s_1) = (4, 3, 1, 0)$ and Figure 2 shows a split tree with cardinality 21 and parameters $(b, s, s_0, s_1) = (2, 4, 0, 2)$.

We can assume that the components V_i of the split vector \mathcal{V} are identically distributed. If this were not the case they can anyway be made identically distributed by using a random permutation, see [3]. Let V be a random variable with this distribution. This gives (because $\sum_i V_i = 1$) that $\mathbf{E}(V) = \frac{1}{b}$. We use the notation T^n to denote a split tree with n balls. However, note that even conditioned on the fact that the split tree has n balls, the number of nodes N , is still a random number. The only parameters that are important

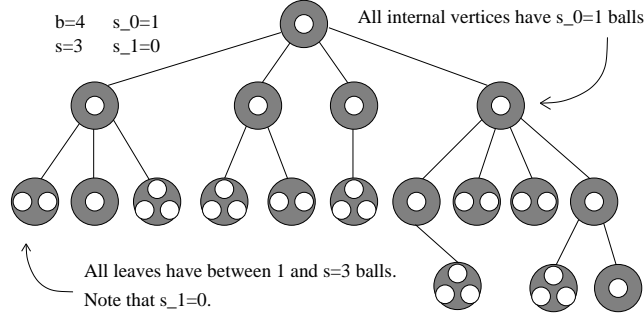


Figure 1: *This figure illustrates a split tree with parameters $b = 4$, $s = 3$, $s_0 = 1$ and $s_1 = 0$.*

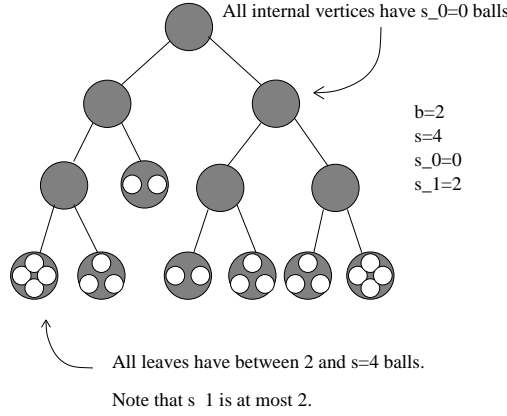


Figure 2: *This figure illustrates a split tree with parameters $b = 2$, $s = 4$, $s_0 = 0$ and $s_1 = 2$.*

in this work (and in general these parameters are the important ones for most results concerning split trees) are the cardinality n , the branch factor b and the split vector \mathcal{V} ; this is illustrated in Section 1.4. As an example, in the binary search tree considered as a split tree, $b = 2$ and the split vector \mathcal{V} is $(U, 1 - U)$ where U is a uniform $U(0, 1)$ random variable. This is a beta $(1, 1)$ random variable. In fact for many important split trees V is beta-distributed. (The other parameters for the binary search tree considered as a split tree are $s = 1$, $s_0 = 1$ and $s_1 = 0$.) For the binary search tree the number of balls n is the same as the number of vertices N ; this is not true for split trees in general.

1.2 Notation

In this section some of the notation that we use in the present study is collected.

Let T^n denote a split tree with n balls; for simplicity we often write T . Let $\mathbf{V}(T)$ denote the set of vertices in a rooted tree T . We write $|S|$ for the number of vertices in a set S . Note that for the number of vertices N we have $N = |\mathbf{V}(T^n)|$. Let T_v be a subtree rooted at v . Let n_v denote the number of balls in the subtree rooted at vertex v and let N_v denote the number of vertices. Note that $N_v = |\mathbf{V}(T_v)|$.

Let D_n denote the depth of the last inserted ball in the tree and D_k the depth of the k -th inserted ball when all n balls have been added. Let D_n^* be the average depth, i.e., $D_n^* = \frac{\sum_{k=1}^n D_k}{n}$. We also use the notation D_k^f for the depth of the node of ball k when it is added to the tree; this could differ from D_k since the ball can move during the splitting process: $D_k^f \leq D_k$. Equivalently, D_k^f is the depth of the last ball in a split tree with k balls. Let $d(v)$ denote the depth (or height) of a vertex v , sometimes we just write d for the depth of v .

Let $p(v)$ denote the parent of a vertex v .

There are at least two different types of total path lengths in a tree T that are of interest: the sum of all depths (distances to the root) of the balls in T , and the sum of all the depths of the vertices in T . We denote the former by $\Psi(T)$ and the latter by $\Upsilon(T)$.

We use the standard notations, $N(\mu, \sigma^2)$ for a normal distribution with expected value μ and variance σ^2 , and $\text{Bin}(m, p)$ for a random variable X with a binomial distribution with parameters m and p . We also use the notation mixed binomial distribution (X, Y) or for short $\text{mBin}(X, Y)$ for a binomial distribution where at least one of the parameters X and Y is a random variable (the other one could be deterministic). Let H_n denote the height of a split tree with n balls.

Let T_{v_i} , $i \in \{1, \dots, b^L\}$ be the subtrees rooted at depth $L = \lfloor \beta \log_b \ln n \rfloor$ for some constant β . For simplicity we just write T_i , $i \in \{1, \dots, b^L\}$ for these. Let $d_w(v) := d(v) - d(w)$, i.e., the depth of a vertex v in the subtree T_w . In particular we write $d_i(v) := d(v) - L$ for the depth of a vertex v in the subtrees T_i , $i \in \{1, \dots, b^L\}$.

Recall that V is a random variable with the distribution of the identically distributed components V_i , $i \in \{1, \dots, b\}$ in the split vector $\mathcal{V} = (V_1, \dots, V_b)$. Let $\Delta = V_S$ be the size biased distribution of (V_1, \dots, V_b) , i.e., given (V_1, \dots, V_b) , let $\Delta = V_j$ with probability V_j , see [3]. Let,

$$c := \mathbf{E}(\Delta) = b\mathbf{E}(V^2),$$

and

$$\begin{aligned}\mu &:= \mathbf{E}\left(-\ln \Delta\right) = b\mathbf{E}\left(-V \ln V\right), \\ \sigma^2 &:= \mathbf{Var}\left(\ln \Delta\right) = b\mathbf{E}\left(V \ln^2 V\right) - \mu^2.\end{aligned}\tag{1}$$

Note that the second equalities of μ and σ imply that they are bounded. Similarly all moments of $-\ln \Delta$ are bounded.

For a given $\epsilon > 0$, we say that a vertex v in T^n is “good” if

$$\mu^{-1} \ln n - \ln^{0.5+\epsilon} n \leq d(v) \leq \mu^{-1} \ln n + \ln^{0.5+\epsilon} n,\tag{2}$$

and “bad” otherwise. We write $\mathbf{V}^*(T^n)$ for the set of good vertices in T^n , and for the number of good vertices we write $N^* := |\mathbf{V}^*(T^n)|$.

We say that $Y_m = o_p(a_m)$ if a_m is a positive number and Y_m is a random variable such that $Y_m/a_m \xrightarrow{p} 0$ as $m \rightarrow \infty$. We use two unusual types of order notation; let a_m be a positive number and Y_m a random variable, by the notation $Y_m := \mathcal{O}_{L^p}(a_m)$ we mean that $(\mathbf{E}(Y_m^p))^{\frac{1}{p}} \leq C a_m$ for some constant C , and by the notation $Y_m := o_{L^p}(a_m)$ we mean that $(\mathbf{E}(Y_m^p))^{\frac{1}{p}}/a_m \rightarrow 0$. We use the notation Ω_j for the σ -field generated by $\{n_v, d(v) \leq j\}$. Finally we write \mathcal{G}_j for the σ -field generated by the \mathcal{V} -vectors for all vertices v with $d(v) \leq j$.

1.3 A weak law and a central limit law for the depth

In [3] Devroye presents a weak law of large numbers and a central limit law for D_n (the depth of the last inserted ball). If $\mathbf{P}(V = 1) = 0$ and $\mathbf{P}(V = 0) < 1$ then

$$\frac{D_n}{\ln n} \xrightarrow{p} \mu^{-1},\tag{3}$$

and

$$\frac{\mathbf{E}(D_n)}{\ln n} \rightarrow \mu^{-1}.\tag{4}$$

From the following lemma it follows easily (as we explain below) that (4) also holds for the average depth D_n^* . Recall that D_k is the depth of the k -th ball in the tree when all n balls are added.

Lemma 1.1. *For $i \leq j$, we have $D_i \leq D_j$ in stochastic sense.*

Proof. We show this by showing that for an arbitrary $i \in \{1, \dots, n-1\}$, $D_i \leq D_{i+1}$, where the inequalities and equalities below are in the stochastic sense only. We show this by the use of coupling arguments.

First consider two identical copies T and \hat{T} of the split tree when $i-1$ balls have been added, where we let \hat{v} in \hat{T} denote the corresponding vertex of v in T . More precisely, we consider two split trees T and \hat{T} with the same split vectors in all vertices of the infinite skeleton tree, and if a ball k , $k \leq i-1$, is added to v in T then ball k is added to \hat{v} in \hat{T} . We now assume that we add the two balls i and $i+1$ to T and \hat{T} .

If ball i and ball $i+1$ are added to different leaves l_1 and l_2 in T then in \hat{T} we let them switch positions, i.e., ball i is added to \hat{l}_2 and ball $i+1$ is added to \hat{l}_1 . (Recall the notation D_k^f from Section 1.2.) Hence, it is obvious for reasons of symmetry that $D_i^f \stackrel{d}{=} D_{i+1}^f$. When the balls $\in \{i+2, \dots, n\}$ are added, we add them to the corresponding vertices in T and \hat{T} . Thus, the two trees are identical in the whole process except for that ball i and ball $i+1$ always have switched positions in T and \hat{T} . Hence, by symmetry $D_i \stackrel{d}{=} D_{i+1}$.

If ball i and ball $i+1$ are added to the same leaf l in T then there are three different cases:

If $n_l \leq s-2$, so that l does not split when also ball i and ball $i+1$ have been added, then T and \hat{T} are still identical since ball i and ball $i+1$ stay in l . When more balls are added we can again assume that ball i and ball $i+1$ have switched positions in T and \hat{T} at every step of the recursive construction until all n balls are added. Hence, by symmetry $D_i \stackrel{d}{=} D_{i+1}$.

If $n_l = s-1$, so that l gets $s+1$ balls when the new balls are added, l splits according to the usual splitting process when ball $i+1$ is added. Again we let ball i and ball $i+1$ switch positions in T and \hat{T} . This means that if ball i is added to v_1 and ball $i+1$ is added to v_2 in T , then in \hat{T} ball i is added to \hat{v}_2 and ball $i+1$ is added to \hat{v}_1 . Thus, again by symmetry $D_i^f \stackrel{d}{=} D_{i+1}^f$. By using the same type of argument as in the cases above we get $D_i \stackrel{d}{=} D_{i+1}$.

If $n_l = s$, so that l in T gets $s+2$ balls when the new balls are added, we let l split according to the usual splitting process where l keeps s_0 balls and sends the other balls to its children.

If ball i is one of the s_0 balls in the children then it is obvious without using the coupling that $D_i^f \leq D_{i+1}^f$ and also $D_i \leq D_{i+1}$.

If ball i is not one of the s_1 balls in the children of l in T and ball i is added to v_1 and ball $i+1$ is added to v_2 , then in \hat{T} we can again assume that ball i is added to \hat{v}_2 and ball $i+1$ is added to \hat{v}_1 . Thus, in the stochastic sense $D_i^f \stackrel{d}{=} D_{i+1}^f$, and $D_i \stackrel{d}{=} D_{i+1}$.

If ball i is one of the s_1 balls in T , we use a related but not an identical

type of coupling argument as in the previous cases. In this case ball i is added by uniformly choosing one of the b children of l each with probability $\frac{1}{b}$, while ball $i + 1$ is added by using the probabilities given by the components in the split vector \mathcal{V}_l of l . Again T and \hat{T} are identical until $i - 1$ balls are added barring the possibility of variation in the split vectors of the vertices above the leaves as described below. If ball i in T goes to a child v_1 of l related to a component V_j in \mathcal{V}_l , then we add ball $i + 1$ in \hat{T} to \hat{v}_1 with probability $\min\{1, \frac{V_j}{1/b}\}$ and to one of the other children related to a component $V_k > 1/b$ with probability $\max\{0, 1 - \frac{V_j}{1/b}\}$, so that the sum of the probabilities gives the right marginal distribution. Assume that ball i is added to the child v of l in T and ball $i + 1$ is added to the child \hat{w} of \hat{l} in \hat{T} . This means that \hat{w} relates to a component of the split vector of \hat{l} at least as large as the component of the split vector of l related to v . Now we can assume that the split vectors in the vertices in the subtree rooted at v correspond to the split vectors in the vertices in the subtree rooted at \hat{w} . This means that we can assume that when ball number j in the subtrees is added it goes to the corresponding vertex in both of the subtrees. However, note that the balls could have different labels if we consider their original label in the whole tree, since the subtree rooted in \hat{w} could have more balls than the subtree rooted in v . Thus, as long as the subtrees have the same number of balls, new balls are added to the corresponding positions in these subtrees, and ball i and ball $i + 1$ are also held by vertices of corresponding positions. This construction shows that if the subtrees rooted in \hat{w} and v have k and l balls, respectively, where $k > l$, and ball i in T_v is in vertex h , then ball $i + 1$ in \hat{T} is in a subtree of $\hat{T}_{\hat{w}}$ with root corresponding to the position of h . This shows that in the stochastic sense $D_i \leq D_{i+1}$.

Hence, in all cases, $D_i \leq D_{i+1}$ stochastically and thus for $i < j$, it follows that $D_i \leq D_j$ in stochastic sense. \square

This means in particular that for all $k \leq n$, $\mathbf{E}(D_k) \leq \mathbf{E}(D_n)$. Since the sum of $\frac{\mathbf{E}(D_k)}{n}$ for $k < \frac{n}{\ln^2 n}$ is $o(1)$, we can ignore the balls $k < \frac{n}{\ln^2 n}$. We consider the balls $k \geq \frac{n}{\ln^2 n}$. For ball $k \geq \frac{n}{\ln^2 n}$ it follows from (4) that $\mathbf{E}(D_k^f) \sim \mu^{-1} \ln n$. Thus, since for all $k \geq \frac{n}{\ln^2 n}$, $\mathbf{E}(D_k^f) \leq \mathbf{E}(D_k) \leq \mathbf{E}(D_n)$, we get

$$\frac{\mathbf{E}(D_n^*)}{\ln n} \rightarrow \mu^{-1}. \quad (5)$$

Furthermore, see [3, Theorem 1], if $\sigma > 0$, and assuming that V is not

monoatomic, i.e., we don't have $\mathcal{V} \equiv \frac{1}{b}$,

$$\frac{D_n - \mu^{-1} \ln n}{\sqrt{\sigma^2 \mu^{-3} \ln n}} \xrightarrow{d} N(0, 1), \quad (6)$$

where $N(0, 1)$ denotes the standard Normal distribution and \xrightarrow{d} denotes convergence in distribution. Tries are special forms of split trees with a random permutation of deterministic components (p_1, p_2, \dots, p_b) and therefore not as random as many other examples. (In the literature tries have also been treated separately to other random trees of logarithmic height.) Of all the most common examples of split trees only some special cases of tries (the symmetric tries and symmetric digital search trees) have a monoatomic distribution of V . From (6) it follows that “most” nodes lie at $\mu^{-1} \ln n + \mathcal{O}(\sqrt{\ln n})$.

1.4 Subtrees

For the split tree where the number of balls $n > s$, there are s_0 balls in the root vertex and the cardinalities of the b subtrees are distributed as (s_1, \dots, s_1) plus a multinomial vector $(n - s_0 - bs_1, V_1, \dots, V_b)$. Thus, conditioning on the random \mathcal{V} -vector that belongs to the root, the subtrees rooted at the children have cardinalities close to nV_1, \dots, nV_b . This is often used in applications of random binary search trees. In particular, we used this fact frequently in [11]. “The split tree generating algorithm” described above, and the fact that a $\text{mBin}(X, p_1)$ in which X is $\text{Bin}(m, p_2)$ is distributed as a $\text{Bin}(m, p_1 p_2)$, give in a stochastic sense, an upper bound on the number of balls n_v in a subtree rooted at a vertex v : Let v be a vertex at depth d , conditioning on \mathcal{G}_d (i.e., the σ -field generated by the \mathcal{V} vectors for all vertices v with $d(v) \leq d$), gives

$$\begin{aligned} n_v &\leq \text{Bin}(n, \prod_{j=1}^d W_{j,v}) + \text{Bin}(s_1, \prod_{j=2}^d W_{j,v}) \\ &\quad + \text{Bin}(s_1, \prod_{j=3}^d W_{j,v}) + \dots + \text{Bin}(s_1, W_{d,v}) + s_1, \end{aligned} \quad (7)$$

where $W_{j,v}, j \in \{1, \dots, d\}$ are i.i.d. random variables given by the split vectors associated with the nodes in the unique path from v to the root. This means in particular that $W_{j,v} \stackrel{d}{=} V$. However, we note that the terms in (7) are not independent. Also observe that \mathcal{G}_d is equivalently the σ -field generated by $W_{j,v}, j \in \{1, \dots, d\}$ for all v with $d(v) = d$. Similarly, we also have a lower

bound for n_v , i.e., for v at depth d , conditioning on \mathcal{G}_d in stochastic sense,

$$\begin{aligned} n_v &\geq \text{Bin}(n, \prod_{j=1}^d W_{j,v}) - \text{Bin}(s, \prod_{j=2}^d W_{j,v}) \\ &\quad - \text{Bin}(s, \prod_{j=3}^d W_{j,v}) - \cdots - \text{Bin}(s, W_{d,v}); \end{aligned} \quad (8)$$

we can replace the term s by $s_0 + bs_1 \leq s$ for a sharper bound. As in (7) the terms in (8) are not independent.

Recall that for a $\text{Bin}(m, p)$ distribution, the expected value is mp and the variance is $mp(1-p)$. Thus, Chebyshev's inequality applied to the dominating term $\text{Bin}(n, \prod_{j=1}^d W_{j,v})$ in (7) gives that n_v for v at depth d is close to

$$M_v^n := nW_{1,v}W_{2,v} \cdots W_{d,v}. \quad (9)$$

More precisely by using (7) and (8), the Chebyshev and Markov inequalities give for v with $d(v) = d$, that for large n ,

$$\begin{aligned} \mathbf{P}\left(\left|n_v - n \prod_{j=1}^d W_{j,v}\right| > n^{0.6}\right) &\leq 4 \frac{\mathbf{E}\left(\mathbf{Var}\left(\text{Bin}(n, \prod_{j=1}^d W_{j,v}) \middle| \mathcal{G}_d\right)\right)}{n^{1.2}} \\ &\quad + 4 \frac{\mathbf{E}\left(\mathbf{E}\left(\text{Bin}(s, \prod_{j=2}^d W_{j,v}) + \text{Bin}(s, \prod_{j=3}^d W_{j,v}) + \cdots + s \middle| \mathcal{G}_d\right)\right)}{n^{0.6}} \\ &\leq \frac{4nb^{-d}}{n^{1.2}} + \frac{\sum_{k=1}^{\infty} 4sb^{-k}}{n^{0.6}} \leq \frac{1}{n^{0.1}}. \end{aligned} \quad (10)$$

Since the n_v 's (conditioned on the split vectors) for all v at the same depth are identically distributed, we sometimes skip the vertex index of $W_{j,v}$ and just write W_j .

1.5 Renewal Theory

Renewal theory is a widely used branch of probability theory that generalizes Poisson processes to arbitrary holding times. A classic in this field is Feller [5] on recurrent events. First we recollect some standard notation. Let $X_0 = 0$ a.s.. Let X_k , $k \geq 1$, be i.i.d. nonnegative random variables distributed as X and let S_m , $m \geq 1$, be the partial sums. Let F denote the distribution function of X , and let F_m be the distribution function of S_m , $m \geq 0$. Thus, for $x \geq 0$,

$$F_0(x) = 1, \quad F_1(x) = F(x), \quad F_m(x) = F^{m*}(x),$$

i.e., F_m equals the m -fold convolution of F itself. The “renewal counting process” $\{\mathcal{N}(t), t \geq 0\}$ is defined by

$$\mathcal{N}(t) := \max\{m : S_m \leq t\},$$

which one can think of as the number of renewals before time t of an object with a lifetime distributed as the random variable X . In the specific case when $X \stackrel{d}{=} \text{Exp}(\lambda)$, $\{\mathcal{N}(t), t \geq 0\}$ is a “Poisson process”. An important well studied function is the so called “standard renewal function” defined as

$$V(t) := \sum_{m=0}^{\infty} F_m(t), \quad (11)$$

which one can easily show is equal to $\mathbf{E}(\mathcal{N}(t))$. The renewal function $V(t)$ satisfies the so called renewal equation

$$V(t) = 1 + (V * dF)(t), \quad t \geq 0.$$

For a broader introduction to renewal theory, see e.g. [1], [6], [7] and [9]. One of the main purposes of this study is to introduce renewal theory in the context of split trees. Recall from (9) in Section 1.4 that the subtree size n_v for v at depth k , is close to $nW_1W_2 \dots W_k$, where W_j , $j \in \{1, \dots, k\}$, are independent random variables distributed as V . Now let $Y_k := -\sum_{j=1}^k \ln W_j$, and for simplicity we also denote the summands $\varpi_r := -\ln W_j$. Note that $nW_1W_2 \dots W_k = ne^{-Y_k}$. Recall that in a binary search tree, the split vector $\mathcal{V} = (V_1, V_2)$ is distributed as $(U, 1 - U)$ where U is a uniform $U(0, 1)$ random variable. For this specific case of a split tree the sum Y_k , (where W_j , $j \in \{1, \dots, k\}$, in this case are i.i.d. uniform $U(0, 1)$ random variables) is distributed as a $\Gamma(k, 1)$ random variable. This fact is used by, e.g., Devroye in [4] to determine the height of a binary search tree. For general split trees there is no simple common distribution function of $\sum_{j=1}^k \ln W_j$, instead renewal theory can be used.

Let

$$\nu_k(t) := b^k \mathbf{P}(Y_k \leq t).$$

We define the renewal function

$$U(t) := \sum_{k=1}^{\infty} \nu_k(t). \quad (12)$$

We also denote $\nu(t) := \nu_1(t) = b \mathbf{P}(\varpi_r \leq t)$. For $U(t)$ we obtain the following renewal equation

$$U(t) = \nu(t) + \sum_{k=1}^{\infty} (\nu_k * d\nu)(t) = \nu(t) + (U * d\nu)(t). \quad (13)$$

2 Main Results

In this section we present the main theorems of this work.

(A1). *In this work we assume as in Section 1.3 that $\mathbf{P}(V = 1) = 0$, and we also assume for simplicity that $\mathbf{P}(V = 0) = 0$ and that $-\ln V$ is non-lattice.*

The reason for the non-lattice assumption (A1) is that we use renewal theory and there it often becomes necessary to distinguish between lattice and non-lattice distributions. Note that the assumption that V is not monoatomic in Section 1.3 is included in the assumption that $-\ln V$ is non-lattice. Again of the common split trees only for some special cases of tries and digital search trees does $-\ln V$ have a lattice distribution. Our first main result is on the relation between the number of vertices N (recall that this is a random variable) and the number of balls n .

Theorem 2.1. *There is a constant α depending on the type of split tree such that*

$$\mathbf{E}(N) = \alpha n + o(n), \quad (14)$$

and

$$\mathbf{Var}(N) = o(n^2). \quad (15)$$

Recall that there is a central limit law for the depth D_n in (6) so that most vertices are close to $\frac{\ln n}{\mu} + \mathcal{O}(\sqrt{\ln n})$, our next result sharpens this fact. Recall that for any constant $\epsilon > 0$, we say that a vertex v in T^n is “good” if

$$\mu^{-1} \ln n - \ln^{0.5+\epsilon} n \leq d(v) \leq \mu^{-1} \ln n + \ln^{0.5+\epsilon} n,$$

and “bad” otherwise.

Theorem 2.2. *For any choice of $\epsilon > 0$, the number of bad nodes in T^n is bounded by $\mathcal{O}_{L^1}\left(\frac{n}{\ln^k n}\right)$ for any constant k .*

In the third main result we sharpen the limit laws in (4) and (5) for the expected value of the depth of the last ball D_n and the average depth D_n^* . We also find the first asymptotic of the variances of the k :th ball D_k for all k , $\frac{n}{\ln n} \leq k \leq n$.

Theorem 2.3. *For the expected value of the depth of the last ball we have*

$$\frac{\mathbf{E}(D_n) - \mu^{-1} \ln n}{\sqrt{\ln n}} \longrightarrow 0, \quad (16)$$

and the same result holds for the average depth D_n^* , i.e.,

$$\frac{\mathbf{E}(D_n^*) - \mu^{-1} \ln n}{\sqrt{\ln n}} \longrightarrow 0. \quad (17)$$

Furthermore, for the variance of the depth of the k :th ball we have that for all $\frac{n}{\ln n} \leq k \leq n$,

$$\frac{\mathbf{Var}(D_k)}{\ln n} \longrightarrow \sigma^2 \mu^{-3}. \quad (18)$$

We complete this section by stating two corollaries of Theorem 2.3. Recall that we write $\mathbf{V}^*(T^n)$ for the set of good vertices in T^n , i.e., those with depths that belong to the strip in (2).

Corollary 2.1. *Summing over all vertices give*

$$\mathbf{E}\left(\sum_{v \in \mathbf{V}^*(T^n)} (d(v) - \mu^{-1} \ln n)^2\right) = \alpha n \mu^{-3} \sigma^2 \ln n + o(n \ln n). \quad (19)$$

For the good vertices we also have

$$\mathbf{E}\left(\sum_{v \in \mathbf{V}^*(T^n)} (d(v) - \mu^{-1} \ln n)^2\right) = \alpha n \mu^{-3} \sigma^2 \ln n + o(n \ln n). \quad (20)$$

We write $\mathbf{V}^*(T_i)$ for the set of good vertices in T_i .

Corollary 2.2. *Let $L = \lfloor \beta \log_b \ln n \rfloor$ for some large constant β . Then, summing over all vertices give*

$$\sum_{i=1}^{b^L} \sum_{v \in \mathbf{V}^*(T_i)} \frac{(d_i(v) - \mu^{-1} \ln n_i)^2}{\mu^{-3} \ln^3 n_i} = \frac{\sigma^2 \alpha n}{\ln^2 n} + o_p\left(\frac{n}{\ln^2 n}\right) \quad (21)$$

and for the good vertices we also have

$$\sum_{i=1}^{b^L} \sum_{v \in \mathbf{V}^*(T_i)} \frac{(d_i(v) - \mu^{-1} \ln n_i)^2}{\mu^{-3} \ln^3 n_i} = \frac{\sigma^2 \alpha n}{\ln^2 n} + o_p\left(\frac{n}{\ln^2 n}\right). \quad (22)$$

3 Some Fundamental Renewal Theory Results

The main goal of this section is to present a renewal theory lemma and a corollary of this lemma, which are both frequently used in this study. In contrast to standard renewal theory the distribution function $\nu(t)$ in (13) is not a probability measure. However, to solve (13) we can apply [1, Theorem VI.5.1] which deals with non probability measures. The result we get is presented in the following lemma.

Lemma 3.1. *The renewal function $U(t)$ in (12) satisfies*

$$U(t) = (\mu^{-1} + o(1))e^t \text{ as } t \rightarrow \infty. \quad (23)$$

Proof. Since the distribution function $\nu(t)$ is not a probability measure, we define another (“conjugate” or “tilted”) measure ω on $[0, \infty)$ by

$$d\omega(t) = e^{-t}d\nu(t).$$

Recall from Section 1.2 that $\Delta = V_S$ is the size biased distribution of (V_1, \dots, V_b) . We note that $\omega(x)$ is the distribution function of the random variable $-\ln \Delta$ since

$$\begin{aligned} \mathbf{P}(-\ln \Delta \leq x) &= \mathbf{E}\left(\mathbf{E}\left(I\{-\ln V_S \leq x\} \middle| (V_1, \dots, V_b)\right)\right) = \\ &= \mathbf{E}\left(\sum_{i=1}^b I\{-\ln V_i \leq x\} V_i\right) = b\mathbf{E}\left(I\{-\ln V \leq x\} e^{-\ln(V)}\right) = \omega(x). \end{aligned}$$

Thus, ω is a probability measure. Further, by recalling $\mu := \mathbf{E}(-\ln \Delta)$ and $\sigma^2 := \mathbf{Var}(-\ln \Delta)$ gives

$$\mathbf{E}(\omega) = \mu, \quad \text{and} \quad \mathbf{Var}(\omega) = \sigma^2. \quad (24)$$

Define $\widehat{U}(t) := e^{-t}U(t)$ and $\widehat{\nu}(t) := e^{-t}\nu(t)$. We shall apply [1, Theorem VI.5.1], but first we need to show that the condition that $\widehat{\nu}(t)$ is “directly Riemann integrable” (d.R.i.) is satisfied. Note that $\widehat{\nu}(t) \leq be^{-t}$, and thus since $\widehat{\nu}(t)$ is also continuous almost everywhere, by [1, Proposition IV.4.1.(iv)] it follows that $\widehat{\nu}(t)$ is d.R.i. if be^{-t} is d.R.i.. That be^{-t} is d.R.i. follows by applying [1, Proposition IV.4.1.(v)], since be^{-t} is a nonincreasing and Lebesgue integrable function. Then by applying [1, Theorem VI.5.1] and (24) we get

$$\widehat{U}(t) = \widehat{\nu}(t) + (\widehat{U} * d\omega)(t), \quad (25)$$

where $\omega(t)$ is a probability measure, and

$$\widehat{U}(t) \rightarrow \mu^{-1} \int_0^\infty \widehat{\nu}(x) dx = \mu^{-1} \int_0^\infty \nu(x) e^{-x} dx =: \kappa. \quad (26)$$

Integration by parts now gives

$$\begin{aligned} \kappa &= \mu^{-1} \left(b \left| -e^{-t} \mathbf{P}(\varpi_r \leq t) \right|_0^\infty - \int_0^\infty -e^{-t} d\nu(t) \right) \\ &= \mu^{-1} b \mathbf{E}(e^{-\varpi_r}) = \mu^{-1}. \end{aligned} \quad (27)$$

Thus, $U(t) = (\mu^{-1} + o(1))e^t$. \square

The following result is a very useful corollary of Lemma 3.1. We write for v at depth $d(v)$, $M_v^n := n \prod_{j=1}^{d(v)} W_j$. Recall from (9) in Section 1.4 that this is close to the real subtree size n_v .

Corollary 3.1. *By taking the sum over vertices v , $d(v) = k$ and letting $\frac{n}{K} \rightarrow \infty$, we get that the expected number of nodes with $M_v^n \geq K$ is equal to*

$$\begin{aligned} \mathbf{E}(\left| v \in V(T^n); M_v^n \geq K \right|) &= \sum_{k=0}^\infty b^k \mathbf{P}(M_v^n \geq K) \\ &=: U(\ln n - \ln K) + 1 = (\mu^{-1} + o(1)) \frac{n}{K}. \end{aligned} \quad (28)$$

Proof. By using Lemma 3.1 we get

$$\begin{aligned} \sum_{d=0}^\infty b^d \mathbf{P}\left(n \prod_{j=1}^d W_{j,v} \geq K\right) &= \sum_{d=0}^\infty b^d \mathbf{P}(Y_d \leq \ln n - \ln K) \\ &= (\mu^{-1} + o(1)) \frac{n}{K}. \end{aligned} \quad (29)$$

\square

We complete this section with a more general result in renewal theory, and a corollary of a more specific result that is valid for the renewal function $U(t)$ in (12).

Theorem 3.1. *Let F be a non-lattice probability measure and suppose that $0 < \mu = \mathbf{E}(X) = \int_0^\infty x dF(x) < \infty$ and $\mathbf{E}(X^2) = \sigma^2 + \mu^2 < \infty$.*

Let

$$Z(t) = z(t) + \int_0^t Z(t-u) dF(u), \quad t \geq 0, \quad (30)$$

where $z(t)$ is a nonnegative function, such that $a := \int_0^\infty z(u)du < \infty$. Define

$$G(x) = \int_0^x \left(Z(t) - \frac{a}{\mu} \right) dt. \quad (31)$$

Then

$$\lim_{x \rightarrow \infty} G(x) = -\frac{1}{\mu} \int_0^\infty uz(u)du + a \frac{\sigma^2 + \mu^2}{2\mu^2}. \quad (32)$$

Proof. Let $V(t)$ be the standard renewal function in (11), where $F_m(t) = \mathbf{P}\left(\sum_{k=0}^m X_k \leq t\right)$. By applying [1, Theorem IV.2.4],

$$Z(t) = \int_0^t z(t-u)dV(u) = \int_0^\infty z(u)dV(t-u), \quad (33)$$

where the last equality follows because $V(t) = 0$ for $t \leq 0$. By applying (33) and Fubini's Theorem we get

$$G(x) = \int_0^\infty z(u) \int_0^x dV(t-u)du - \frac{ax}{\mu} = \int_0^\infty z(u)V(x-u)du - \frac{ax}{\mu}. \quad (34)$$

Hence,

$$\begin{aligned} G(x) &= \int_0^\infty z(u) \left(V(x-u) - \frac{x}{\mu} \right) du \\ &= -\frac{1}{\mu} \int_0^x z(u)u du - \frac{1}{\mu} \int_x^\infty z(u)x du + \int_0^x z(u) \left(V(x-u) - \frac{x-u}{\mu} \right) du. \end{aligned} \quad (35)$$

From [1, Proposition VI.4.1] we have $V(t) - \frac{t}{\mu} \rightarrow \frac{\sigma^2 + \mu^2}{2\mu^2}$ and by [1, Proposition VI.4.2], $0 \leq V(t) - \frac{t}{\mu} \leq \frac{\sigma^2 + \mu^2}{\mu^2}$. Hence, the Lebesgue dominated convergence theorem applied to the last integral in (35) gives

$$\lim_{x \rightarrow \infty} \int_0^\infty z(u) \left(V(x-u) - \frac{x-u}{\mu} \right) I\{u \leq x\} du = \int_0^\infty z(u) \frac{\sigma^2 + \mu^2}{2\mu^2} du. \quad (36)$$

Note that for all x , $\int_x^\infty z(u)(u-x)du \geq 0$. Thus, if $\int_0^\infty z(u)u du$ is integrable $\lim_{x \rightarrow \infty} \int_x^\infty z(u)x du = 0$, and the convergence result in (32) obviously follows. If $\int_0^\infty z(u)u du$ is not integrable then we have a special case of (32), i.e., $\lim_{x \rightarrow \infty} G(x) = -\infty$. \square

We define the function

$$W(x) = \int_0^x e^{-t}(U(t) - \mu^{-1}e^t)dt. \quad (37)$$

The next result is a corollary of Theorem 3.1.

Corollary 3.2. *The function $W(x)$ in (37) satisfies*

$$W(x) = \frac{\sigma^2 - \mu^2}{2\mu^2} - \mu^{-1} + o(1) \quad \text{as } x \rightarrow \infty. \quad (38)$$

Proof. We apply Theorem 3.1 to $Z(t) = \widehat{U}(t) = e^{-t}U(t)$ defined in the proof of Lemma 3.1 (recall that $\widehat{U}(t)$ satisfies the renewal equation in (25)). Now, the constant a as defined in Theorem 3.1, satisfies $a = \int_0^\infty \widehat{\nu}(u)du$, thus from (26) and (27) we get $a = 1$. Using (24) and (26)–(27) gives,

$$\int_0^\infty \widehat{\nu}(u)udu = \int_0^\infty e^{-u}\nu(u)udu = \int_0^\infty e^{-u}\nu(u)du + \int_0^\infty ue^{-u}d\nu(u) = 1 + \mu.$$

□

4 Proofs of the Main Results

4.1 Proof of Theorem 2.1

4.1.1 Lemmas of Theorem 2.1

We present below some crucial lemmas by which we can then prove Theorem 2.1. The proofs of these lemmas are given in Section 4.1.4 below. The first lemma is fundamental for the proof.

Lemma 4.1. *For the first moment of the number of vertices N we have*

$$\mathbf{E}(N) = \mathcal{O}(n) \quad (39)$$

and for the second moment of N we have

$$\mathbf{E}(N^2) = \mathcal{O}(n^2). \quad (40)$$

Lemma 4.2. *Adding K balls to a tree will only affect the expected number of nodes in a split tree by $\mathcal{O}(K)$ nodes.*

Let R be the set of vertices such that conditioned on the split vectors, $r \in R$, if $M_r^n := n \prod_{j=1}^{d(r)} W_j < B$ and $M_{p(r)}^n := n \prod_{j=1}^{d(r)-1} W_j \geq B$, recall that $p(r)$ is the parent of r . For now we just let B be large; however, later our choice of B will be more precise. To show (14) we consider all subtrees rooted at some vertex $r \in R$. We denote these subtrees by $T_{r,B}$, $r \in R$. Recall from (9) that with “large” probability the cardinality n_r is “close” to M_r^n . We will show that in fact we can replace n_r by M_r^n in our calculations. Let n_r be the number of balls and let N_r be the number of nodes in the $T_{r,B}$ subtree. Corollary 3.1 implies that most vertices are in the $T_{r,B}$ subtrees, i.e.,

$$\mathbf{E}(N) = \mathbf{E}\left(\sum_{r \in R} N_r\right) + \mathcal{O}\left(\frac{n}{B}\right). \quad (41)$$

The next lemma shows that the expected number of vertices in the $T_{r,B}$ subtrees with subtree sizes n_r that differ significantly from M_r^n is bounded by a “small” error term for large B . Since the variance of a $\text{Bin}(m, p)$ distribution is $m(p - p^2)$, the Chebyshev and Markov inequalities give similarly as in (10) that for large B ,

$$\mathbf{P}\left(|n_r - M_r^n| \geq B^{0.6}\right) \leq 4 \frac{\mathbf{E}(M_r^n)}{B^{1.2}} + \frac{\sum_{k=1}^{\infty} 4sb^{-k}}{B^{0.6}} \leq \frac{1}{B^{0.1}}. \quad (42)$$

From (41) we have

$$\begin{aligned} \mathbf{E}(N) &= \mathbf{E}\left(\sum_{r \in R} N_r I\{|n_r - M_r^n| \geq B^{0.6}\}\right) + \\ &\quad \mathbf{E}\left(\sum_{r \in R} N_r I\{|n_r - M_r^n| \leq B^{0.6}\}\right) + \mathcal{O}\left(\frac{n}{B}\right). \end{aligned} \quad (43)$$

Lemma 4.3. *The expected value of the number of nodes that are not in the $T_{r,B}$, $r \in R$, subtrees with subtree size n_r that differs from M_r^n with at least $B^{0.6}$ balls, is*

$$\mathbf{E}\left(\sum_{r \in R} N_r I\{|n_r - M_r^n| \geq B^{0.6}\}\right) = \mathcal{O}\left(\frac{n}{B^{0.1}}\right), \quad (44)$$

hence, from (43)

$$\mathbf{E}(N) = \mathbf{E}\left(\sum_{r \in R} N_r I\{|n_r - M_r^n| \leq B^{0.6}\}\right) + \mathcal{O}\left(\frac{n}{B^{0.1}}\right). \quad (45)$$

We also sub-divide the $T_{r,B}$, $r \in R$, subtrees into smaller classes, wherein the M_r^n 's in each class are close to each-other. Choose $\gamma := \epsilon^2$ and let $Z := \{B, B - \gamma B, B - 2\gamma B, \dots, \epsilon B\}$, where $\epsilon = \frac{1}{k}$ for some positive integer k . We write $R_z \subseteq R$, $z \in Z$, for the set of vertices $r \in R$, such that $M_r^n \in [z - \gamma B, z)$ and $M_{p(r)}^n \geq B$. (Note that the intervals are of length γB and that the set Z contains at most $\frac{1}{\gamma}$ elements.) We write $|R_z|$ for the number of nodes in R_z . The next lemma is a result that we get by the use of renewal theory applied to the renewal function $U(t)$ in (12).

Lemma 4.4. *Let $S := \{1, 1 - \gamma, 1 - 2\gamma, \dots, \epsilon\}$, where $\gamma = \epsilon^2$. Choose $\alpha \in S$ and let $\frac{n}{B} \rightarrow \infty$, then*

$$\frac{\mathbf{E}(|R_{\alpha B}|)}{\frac{n}{B}} = c_\alpha + o(1), \quad (46)$$

for a constant c_α (only depending on α), and also $\sum_{\alpha \in S} c_\alpha = \mathcal{O}(1)$, where the constant in \mathcal{O} is not depending on ϵ .

Before proving these lemmas we show how their use leads to the proof of Theorem 2.1.

4.1.2 Proof of (14) in Theorem 2.1

Proof. For showing (14) it is enough to show that for two arbitrary values of the cardinality n and \hat{n} , where $\hat{n} \geq n$, we have

$$\left| \frac{\mathbf{E}(N)}{n} - \frac{\mathbf{E}(\hat{N})}{\hat{n}} \right| = \mathcal{O}(\epsilon) \quad \text{as } n \rightarrow \infty, \quad \forall \epsilon > 0. \quad (47)$$

Since (47) implies that $\frac{\mathbf{E}(N)}{n}$ is Cauchy it follows that $\frac{\mathbf{E}(N)}{n}$ converges to some constant α as n tends to infinity; hence, we deduce (14).

We will now prove (47).

Recall from Section 4.1.1 that we will consider the subtrees $T_{r,B}$, $r \in R$, rooted at r ; these are defined such that $M_r^n := n \prod_{j=1}^{d(r)} W_j < B$ and $M_{p(r)}^n := n \prod_{j=1}^{d(r)-1} W_j \geq B$.

Let $R' \subseteq R$ be the set of vertices such that $r \in R'$ if

$$|n_r - M_r^n| \leq B^{0.6}. \quad (48)$$

Lemma 4.3 shows that we only need to consider the vertices in $r \in R'$.

Let $R'' \subseteq R'$ be the set of vertices such that $r \in R''$ if $r \in R'$ and

$$\epsilon B < M_r^n < B. \quad (49)$$

We will now explain that it is enough to consider the vertices $r \in R''$.

Corollary 3.1 for $K = B$ gives that the expected number of parents $p(r)$ such that $M_{p(r)}^n \geq B$ is $\mathcal{O}\left(\frac{n}{B}\right)$; thus, since they only have b children each, also the expectation of $|R|$ is $\mathcal{O}\left(\frac{n}{B}\right)$. Hence, for $r \in R'$ by using (39) in Lemma 4.1, we get that the expected number of nodes in the $T_{r,B}$, $r \in R'$, with $M_r^n \leq \epsilon B$ is bounded by $\mathcal{O}(\epsilon n)$.

From (45) in Lemma 4.3, we get

$$\mathbf{E}(N) = \mathbf{E}\left(\sum_{r \in R''} N_r\right) + \mathcal{O}(\epsilon n) + \mathcal{O}\left(\frac{n}{B^{0.1}}\right). \quad (50)$$

Recall that we sub-divide the $T_{r,B}$, $r \in R$, subtrees into smaller classes, wherein the M_r^n 's in each class are close to each-other, by introducing the subsets $R_z \subseteq R$, $z \in Z$, where $Z = \{B, B - \gamma B, B - 2\gamma B, \dots, \epsilon B\}$. Hence, (50) gives

$$\mathbf{E}(N) = \mathbf{E}\left(\sum_{z \in Z} \sum_{r \in R' \cap R_z} N_r\right) + \mathcal{O}(\epsilon n) + \mathcal{O}\left(\frac{n}{B^{0.1}}\right). \quad (51)$$

We will now apply Lemma 4.2 to calculate the expected value in (51). Let r_z be an arbitrarily chosen node in $R' \cap R_z$, where $z \in Z$. By using (48) and Lemma 4.2, for any node $r_z \in R' \cap R_z$, we get that the expected number of nodes in a tree with the number of balls in an interval $[z - \gamma B, z)$ is equal to $\mathbf{E}(N_{r_z}) + \mathcal{O}(\gamma B)$. By using (51) this implies that

$$\mathbf{E}(N) = \sum_{z \in Z} \mathbf{E}(|R' \cap R_z|) \left(\mathbf{E}(N_{r_z}) + \mathcal{O}(\gamma B) \right) + \mathcal{O}(\epsilon n) + \mathcal{O}\left(\frac{n}{B^{0.1}}\right). \quad (52)$$

Define a_x as the quotient of the expected number of vertices in a tree with cardinality $\lfloor x \rfloor$ divided by $\lfloor x \rfloor$. Note from Lemma 4.1 that $a_x = \mathcal{O}(1)$.

Recall from Lemma 4.4 that $S = \{1, 1 - \gamma, 1 - 2\gamma, \dots, \epsilon\}$. By using (46) in Lemma 4.4 and applying (42) we have that for each choice of γ and $\alpha \in S$, there is a σ_γ such that for a constant c_α (depending on α),

$$\left| \frac{\mathbf{E}(|R' \cap R_{\alpha B}|)}{\frac{n}{B}} - c_\alpha \right| \leq \gamma^2 + \mathcal{O}\left(\frac{1}{B^{0.1}}\right), \quad (53)$$

whenever $\frac{n}{B} \geq \frac{1}{\sigma_\gamma}$. We now choose $B = \ln n$, where n is the smallest of the two arbitrary values we start with (i.e., $\hat{n} \geq n$). Thus, we have by the choice of B (for n large enough) that $\frac{n}{B} \geq \frac{1}{\sigma_\gamma}$ so that (53) holds.

Note that since $\sum_{\alpha \in S} c_\alpha = \mathcal{O}(1)$, we have that $\sum_{\alpha \in S} c_\alpha \frac{\mathcal{O}(B\gamma)}{B} = \mathcal{O}(\gamma)$. Recall that $\gamma := \epsilon^2$. Thus, for a constant c_α (depending on α) and $a_{\alpha B} = \mathcal{O}(1)$, we get from (52) and (53) that

$$\begin{aligned} \mathbf{E}(N) &= n \sum_{\alpha \in S} c_\alpha \frac{1}{B} \left(\alpha B a_{\alpha B} + \mathcal{O}(B\gamma) \right) + n \sum_{\alpha \in S} \mathcal{O}(a_{\alpha B} \gamma^2) + \mathcal{O}(\epsilon n) = \\ &= n \sum_{\alpha \in S} \alpha a_{\alpha B} c_\alpha + \mathcal{O}(\epsilon n). \end{aligned} \quad (54)$$

In analogy we also get for $\hat{n} \geq n$,

$$\mathbf{E}(\hat{N}) = \hat{n} \sum_{\alpha \in S} \alpha a_{\alpha B} c_\alpha + \mathcal{O}(\epsilon \hat{n}). \quad (55)$$

Thus, (47) follows, which shows (14). □

4.1.3 Proof of (15) in Theorem 2.1

Proof. First note that (40) in Lemma 4.1 implies that $\mathbf{Var}(N) = \mathcal{O}(n^2)$. The purpose is to use the variance formula

$$\mathbf{Var}(Y) = \mathbf{E}(\mathbf{Var}(Y|\mathcal{G})) + \mathbf{Var}(\mathbf{E}(Y|\mathcal{G})), \quad (56)$$

where Y is a random variable and \mathcal{G} is a sub- σ -field, see e.g. [10, exercise 10.17-2]. We consider the subtrees T_i , $1 \leq i \leq b^D$ at depth $D = c \ln n$, choosing the constant c small enough so that the number of nodes Z_D between depth D and the root is $\mathcal{O}(n^\epsilon)$ for some arbitrary small ϵ . Let n_i be the number of balls and N_i the number of nodes in T_i . Conditioned on Ω_D , N_i , $1 \leq i \leq b^D$, are independent and it follows that,

$$\mathbf{Var}(N|\Omega_D) = \mathbf{Var}\left(\sum_{i=1}^{b^D} N_i + Z_D|\Omega_D\right) = \sum_{i=1}^{b^D} \mathbf{Var}(N_i|\Omega_D) = \sum_{i=1}^{b^D} \mathcal{O}(n_i^2). \quad (57)$$

Taking expectation in (57) gives

$$\mathbf{E}(\mathbf{Var}(N|\Omega_D)) = \sum_{i=1}^{b^D} \mathcal{O}(\mathbf{E}(n_i^2)). \quad (58)$$

Recall that Ω_D is the σ -field generated by $\{n_v, d(v) \leq D\}$.

Lemma 4.5. *For $D = c \ln n$ there is a $\delta > 0$, such that*

$$\mathbf{E}\left(\sum_{i=1}^{b^D} n_i^2\right) = \mathcal{O}(n^{2-\delta}). \quad (59)$$

Proof. The representation of subtree sizes in split trees described in (7) in Section 1.4 gives in particular that conditioning on \mathcal{G}_D , n_i for i at depth D is bounded from above (in stochastic sense) by

$$\text{Bin}(n, \prod_{j=1}^D W_j) + s_1 D, \quad (60)$$

where $W_j, j \in \{1, \dots, D\}$, are i.i.d. random variables distributed as V . The fact that the second moment of a $\text{Bin}(m, p)$ is $m^2 p^2 + mp - mp^2$ and the bound of n_i in (60) give

$$\mathbf{E}(n_i^2 | \mathcal{G}_D) \leq n^2 \prod_{j=1}^D W_j^2 + \mathcal{O}(nD \prod_{j=1}^D W_j) + \mathcal{O}(D^2).$$

Note that $\mathbf{E}(W_j^2) < \mathbf{E}(W_j) = \frac{1}{b}$, since $W_j \in (0, 1)$. Hence, there is an $\epsilon > 0$ such that

$$\begin{aligned} \mathbf{E}(n_i^2) &\leq n^2 \prod_{j=1}^D \mathbf{E}(W_j^2) + \mathcal{O}\left(\frac{nD}{b^D}\right) + \mathcal{O}(D^2) \\ &\leq \frac{n^2}{(b + \epsilon)^D} + \mathcal{O}\left(\frac{nD}{b^D}\right) + \mathcal{O}(D^2), \end{aligned} \quad (61)$$

and thus there is a $\delta > 0$ such that

$$\mathbf{E}\left(\sum_{i=1}^{b^D} n_i^2\right) = \mathcal{O}(n^{2-\delta}),$$

which shows (59). □

Thus, (58) and (59) in Lemma 4.5 give

$$\mathbf{E}\left(\mathbf{Var}(N | \Omega_D)\right) = \mathcal{O}(n^{2-\delta}). \quad (62)$$

By applying (14) in Theorem 2.1 gives

$$\mathbf{E}(N | \Omega_D) = \sum_{i=1}^{b^D} \left(\alpha n_i + o(n_i)\right) + \mathbf{E}(Z_D | \Omega_D). \quad (63)$$

Applying (63) gives

$$\mathbf{Var}\left(\mathbf{E}(N|\Omega_D)\right) = \mathbf{Var}\left(\alpha n + o(n)\right) = o(n^2). \quad (64)$$

Thus, by applying the variance formula in (56) we get from (62) and (64) that $\mathbf{Var}(N) = o(n^2)$. \square

Remark 4.1. The proof shows that if we can improve the result in (14) in Theorem 2.1 such that $\mathbf{E}(N) = \alpha n + \mathcal{O}(n^{1-c_1})$ for some constant $c_1 > 0$, we will also get a sharper result for the variance, i.e., $\mathbf{Var}(N) = o(n^{2-c_2})$ for some constant $c_2 > 0$.

4.1.4 Proofs of the Lemmas of Theorem 2.1

Proof of Lemma 4.1. (Note that if $s_0 > 0$ it is always true that $N \leq n$ and if $s_1 > 0$ we always have $N \leq 2n$.) For $s_0 = s_1 = 0$ we can argue as follows: When a new ball is added to the tree the expected number of additional nodes is bounded by the expected number of nodes one gets from a splitting node. Let Z be the number of nodes that one gets when a node of $s+1$ balls splits. Then

$$\mathbf{E}(Z) = \sum_{k=1}^{\infty} k \mathbf{P}(Z = k). \quad (65)$$

Note that once a node gives balls to at least 2 children the splitting process ends. Thus,

$$\mathbf{P}(Z = k | \mathcal{G}_k) = \mathcal{O}\left(\sum_{v, d(v)=k} \prod_{j=1}^k W_{j,v}^{s+1}\right).$$

Hence, (65) implies,

$$\mathbf{E}(Z) = \sum_{k=1}^{\infty} k \mathcal{O}\left((b \mathbf{E}(V^{s+1}))^k\right). \quad (66)$$

There is a $\delta > 0$ such that $b \mathbf{E}(V^{s+1}) \leq b^{-\delta}$ since

$$\mathbf{E}(V^{s+1}) < \mathbf{E}(V) = \frac{1}{b}, \quad (67)$$

for $V \in (0, 1)$. Thus, (66) gives

$$\mathbf{E}(Z) = \sum_{k=1}^{\infty} k \mathcal{O}\left(b^{-k\delta}\right) = \mathcal{O}\left(\frac{b^{-\delta}}{(1 - b^{-\delta})^2}\right) = \mathcal{O}(1). \quad (68)$$

This shows (39).

Now we show (40). Note that (40) obviously holds if $s_1 > 0$ or $s_0 > 0$, since then $N \leq 2n$. Recall that Z is the number of nodes that one gets when a node of $s + 1$ balls splits. Then by the well-known Minkowski's inequality

$$\mathbf{E}(N^2) \leq n^2 \mathbf{E}(Z^2). \quad (69)$$

By similar calculations as in (66)–(68) we get that for some constant $\delta > 0$,

$$\mathbf{E}(Z^2) \leq \sum_{k=1}^{\infty} k^2 \mathbf{P}(Z = k) = \sum_{k=1}^{\infty} k^2 \mathcal{O}(b^{-k\delta}) = \mathcal{O}(1). \quad (70)$$

Thus, (40) follows from (69) and (70). \square

Proof of Lemma 4.1. The proof of this lemma is in analogy with the proof of (39) in Lemma 4.1. Adding one ball to the tree will only increase the vertices if it is added to a leaf with s balls. Recall that Z is the number of nodes that one gets when a node of $s + 1$ balls splits. Hence, (66) gives $\mathbf{E}(Z) = \mathcal{O}(1)$, implying that K balls can only create $\mathcal{O}(K)$ additional nodes. \square

Proof of Lemma 4.3. By applying (42) we get that with probability at least $1 - \frac{1}{B^{0.1}}$,

$$|n_r - M_r^n| \leq B^{0.6}. \quad (71)$$

We have

$$\mathbf{E}\left(\sum_r n_r I\{|n_r - M_r^n| \geq B^{0.6}\}\right) = E_1 + E_2, \quad (72)$$

where

$$\begin{aligned} E_1 &= \mathbf{E}\left(\sum_r n_r I\{|n_r - M_r^n| \geq B^{0.6}\} I\{n_r \leq 2M_r^n\}\right), \\ E_2 &= \mathbf{E}\left(\sum_r n_r I\{|n_r - M_r^n| \geq B^{0.6}\} I\{n_r > 2M_r^n\}\right). \end{aligned}$$

Hence, the facts that $\sum_r M_r^n = \mathcal{O}(n)$ and that the bound in (71) holds with probability $1 - \frac{1}{B^{0.1}}$, give

$$E_1 \leq \mathbf{E}\left(\sum_r 2M_r^n I\{|n_r - M_r^n| \geq B^{0.6}\}\right) = \mathcal{O}\left(\frac{n}{B^{0.1}}\right).$$

Recall that R is the set of vertices such that $r \in R$, if r is the root of a $T_{r,B}$ subtree. We obviously have

$$E_2 \leq \mathbf{E} \left(\sum_v 2(n_v - M_v^n) I\{n_v > 2M_v^n\} I\{v \in R\} \right).$$

By summing over vertices v at depth k we get

$$E_2 \leq \sum_{k=0}^{\infty} 2b^k \mathbf{E} \left((n_v - M_v^n) I\{n_v > 2M_v^n\} \right) \mathbf{P}(v \in R). \quad (73)$$

We write F for the expected value in (73), i.e.,

$$F := \mathbf{E} \left((n_v - M_v^n) I\{n_v > 2M_v^n\} \right).$$

Hence, the conditional Cauchy-Schwarz and the conditional Markov inequalities give

$$\begin{aligned} F &\leq \mathbf{E} \left(\sqrt{\mathbf{E} \left((n_v - M_v^n)^2 \middle| \mathcal{G}_d \right)} \sqrt{\mathbf{P}(n_v > 2M_v^n \middle| \mathcal{G}_d)} \right) \\ &\leq \min \left(\mathbf{E} \left(\frac{\mathbf{E} \left((n_v - M_v^n)^2 \middle| \mathcal{G}_d \right)}{M_v^n} \right), \mathbf{E} \left(\sqrt{\mathbf{E} \left((n_v - M_v^n)^2 \middle| \mathcal{G}_d \right)} \right) \right). \end{aligned} \quad (74)$$

From (7) we have that for all v with $d(v) = d$, conditioned on \mathcal{G}_d (i.e., the σ -field generated by $W_{j,v}, j \in \{1, \dots, d\}$), $n_v \leq n'_v + n''_v$, where

$$\begin{aligned} n'_v &:= \text{Bin}(n, \prod_{j=1}^d W_{j,v}), \\ n''_v &:= \text{Bin}(s_1, \prod_{j=2}^d W_{j,v}) + \dots + \text{Bin}(s_1, W_{d,v}) + s_1. \end{aligned}$$

Thus, (74) gives for $M_v^n \geq 1$,

$$\begin{aligned}
F &\leq \mathbf{E} \left(\frac{\mathbf{E} \left((n_v - M_v^n)^2 \middle| \mathcal{G}_d \right)}{M_v^n} \right) \\
&\leq \mathbf{E} \left(\frac{\mathbf{E} \left((n'_v - M_v^n)^2 \middle| \mathcal{G}_d \right)}{M_v^n} + \frac{\mathbf{E} \left((n''_v)^2 + 2n'_v n''_v - 2n''_v M_v^n \middle| \mathcal{G}_d \right)}{M_v^n} \right) \\
&\leq \mathbf{E} \left(\frac{\mathbf{E} \left((n'_v - M_v^n)^2 \middle| \mathcal{G}_d \right)}{M_v^n} \right) + \mathbf{E} \left((n''_v)^2 \right) + 2\mathbf{E} \left(n''_v \right), \tag{75}
\end{aligned}$$

where we in the last equality apply that $\mathbf{E}(n'_v) = M_v^n$. For $M_v^n < 1$ we apply that (74) gives

$$F \leq \mathbf{E} \left(\sqrt{\mathbf{E} \left((n_v - M_v^n)^2 \middle| \mathcal{G}_d \right)} \right). \tag{76}$$

By applying the fact that the variance of a $\text{Bin}(m, p)$ distribution is $m(p-p^2)$ we get $\mathbf{E} \left((n'_v - M_v^n)^2 \middle| \mathcal{G}_d \right) \leq M_v^n$, and from the Minkowski's inequality we easily deduce that $\mathbf{E} \left((n''_v)^2 \right) = \mathcal{O}(1)$. Hence, by using that we can bound F as in (75) for $M_v^n \geq 1$, and by the bound in (76) for $M_v^n < 1$, we get that $F = \mathcal{O}(1)$. Thus, from (73) we get

$$E_2 \leq \sum_{k=0}^{\infty} b^k \mathcal{O}(1) \mathbf{P}(v \in R). \tag{77}$$

Note that $v \in R$ only if $M_{p(v)}^n \geq B$. Hence, by applying Corollary 3.1 for $K = B$, we get from (77) that $E_2 = \mathcal{O}(\frac{n}{B})$. By applying Lemma 4.1 in combination with (72) and using the bounds of E_1 and E_2 we get

$$\begin{aligned}
\mathbf{E} \left(\sum_r N_r I \{ |n_r - M_r^n| \geq B^{0.6} \} \right) &= \mathcal{O} \left(\mathbf{E} \left(\sum_r n_r I \{ |n_r - M_r^n| \geq B^{0.6} \} \right) \right) \\
&= \mathcal{O} \left(\frac{n}{B^{0.1}} \right). \tag{78}
\end{aligned}$$

□

Proof of Lemma 4.4. Recall the definition of $Y_k = -\sum_{j=1}^k \ln W_j$ and $\nu(t) = b\mathbf{P}(-\ln W_j \leq t)$. Also recall that we write $S = \{1, 1 - \gamma, 1 - 2\gamma, \dots, \epsilon\}$ for $\gamma = \epsilon^2$. We have for $\alpha \in S$

$$\begin{aligned} \mathbf{E}(|R_{\alpha B}|) &= \sum_{k=0}^{\infty} b^{k+1} \left(\mathbf{P}\left(\{Y_k - \ln W_{k+1} > \ln \frac{n}{B} - \ln \alpha\} \cap \{Y_k \leq \ln \frac{n}{B}\}\right) \right. \\ &\quad \left. - \mathbf{P}\left(\{Y_k - \ln W_{k+1} > \ln \frac{n}{B} - \ln(\alpha - \gamma)\} \cap \{Y_k \leq \ln \frac{n}{B}\}\right) \right). \end{aligned}$$

We write $q := \ln \frac{n}{B}$. From the definition of $U(t)$ we have that $\mathbf{E}(|R_{\alpha B}|)$ is equal to

$$\begin{aligned} Z(q) &:= \int_0^q b \left(\mathbf{P}\left(-\ln W_{k+1} > q - t - \ln \alpha\right) \right. \\ &\quad \left. - \mathbf{P}\left(-\ln W_{k+1} > q - t - \ln(\alpha - \gamma)\right) \right) dU(t). \end{aligned}$$

Hence,

$$Z(q) := \int_0^q b \mathbf{P}\left(q - t - \ln \alpha < -\ln W_{k+1} \leq q - t - \ln(\alpha - \gamma)\right) dU(t). \quad (79)$$

We write

$$G(t) := b \mathbf{P}\left(t - \ln \alpha < -\ln W_{k+1} \leq t - \ln(\alpha - \gamma)\right).$$

Thus,

$$Z(q) = (G * dU)(q).$$

Recall that we write $d\omega(t) = e^{-t} d\nu(t)$ where $\omega(t)$ is a probability measure. Recall from (25) that we have

$$\widehat{U}(t) = \widehat{\nu}(t) + (\widehat{U} * d\omega)(t),$$

where $\widehat{U}(t) := e^{-t} U(t)$ and $\widehat{\nu}(t) := e^{-t} \nu(t)$. Thus, by using [1, Theorem VI.5.1] we have for $\widehat{Z}(x) = e^{-x} Z(x)$ and $\widehat{G}(x) = e^{-x} G(x)$ that

$$\widehat{Z}(q) = (\widehat{G} * d\omega)(q).$$

By using (79) this implies that

$$\widehat{Z}(q) = \int_0^q b e^{t-q} \mathbf{P}\left(q - t - \ln \alpha < -\ln W_{k+1} \leq q - t - \ln(\alpha - \gamma)\right) d\omega(t). \quad (80)$$

By using the key renewal theorem [9, Theorem II.4.3] applied to $\widehat{U}(t)$ we get

$$\lim_{q \rightarrow \infty} \widehat{Z}(q) = \frac{b}{\mu} \int_0^\infty e^{-t} \mathbf{P}\left(t - \ln \alpha < -\ln W_{k+1} \leq t - \ln(\alpha - \gamma)\right) dt. \quad (81)$$

Note that $\lim_{q \rightarrow \infty} \widehat{Z}(q) := c_\alpha$, for some constant c_α only depending on α .

Thus, by using $\widehat{Z}(x) = e^{-x} Z(x)$ we get that

$$\mathbf{E}(|R_{\alpha B}|) = \frac{n}{B} c_\alpha + o\left(\frac{n}{B}\right),$$

for the constant c_α (only depending on α), which shows (46).

Also note that we have

$$\sum_{\alpha \in S} c_\alpha = \frac{b}{\mu} \int_0^\infty e^{-t} \sum_{\alpha \in S} \mathbf{P}\left(t - \ln \alpha < -\ln W_{k+1} \leq t - \ln(\alpha - \gamma)\right) dt \quad (82)$$

$$= \frac{b}{\mu} \int_0^\infty e^{-t} \mathbf{P}\left(t < -\ln W_{k+1} \leq t - \ln(\epsilon - \gamma)\right) dt \leq \frac{b}{\mu}. \quad (83)$$

□

4.2 Proof of Theorem 2.2

Proof. We use large deviations to show this theorem (in fact we get a sharper bound of the number of bad nodes). Note that a vertex v belongs to the tree if and only if $n_v \geq 1$. Recall that there is an upper bound of n_v with $d(v) = d$ in (7) above, i.e., conditioning on \mathcal{G}_d in stochastic sense,

$$\begin{aligned} n_v &\leq \text{Bin}\left(n, \prod_{j=1}^d W_{j,v}\right) + \text{Bin}\left(s_1, \prod_{j=2}^d W_{j,v}\right) \\ &\quad + \text{Bin}\left(s_1, \prod_{j=3}^d W_{j,v}\right) + \cdots + \text{Bin}(s_1, W_{d,v}) + s_1, \end{aligned} \quad (84)$$

where $W_{j,v}$, $j \in \{1, \dots, d\}$, are i.i.d. random variables distributed as V . It is enough to just consider the first term $\text{Bin}(n, \prod_{j=1}^d W_{j,v})$ in (84), and prove that the number of bad nodes with $\text{Bin}(n, \prod_{j=1}^d W_{j,v}) \geq 1$ is bounded by $\mathcal{O}_{L^1}\left(\frac{n}{\ln^{k+1} n}\right)$, where we choose k large enough. If $s_1 = 0$, $\text{Bin}(n, \prod_{j=1}^d W_{j,v})$ is the only term in (84). We now explain the fact that we can ignore the terms in n_v that occurs because of the parameter s_1 . Assume that for split

trees with $s_1 = 0$, the number of bad nodes is bounded by $\mathcal{O}_{L^1}\left(\frac{n}{\ln^{k+1}n}\right)$. We first consider the vertices with $d \leq \mu^{-1} \ln n - \ln^{0.5+\epsilon} n$. If $s_1 > 0$, we assume that we first add the n balls as in the construction of a split tree with the parameter $s_1 = 0$. Hence, the number of vertices v with $d \leq \mu^{-1} \ln n - \ln^{0.5+\epsilon} n$, is bounded by $\mathcal{O}_{L^1}\left(\frac{n}{\ln^{k+1}n}\right)$. We now repay the subtree sizes for their potential loss of balls because of $s_1 > 0$. A vertex v at depth d can at most have a loss of $s_1 d$ balls in the subtree rooted at v . These balls cannot give more than $s_1 b d$ nodes to the tree (since only if $s_0 = s_1 = 0$ it is possible for an increment of more than b nodes when a new ball is added to the tree). Thus, since $d \leq \mu^{-1} \ln n$ and the fact that we assume that we have $\mathcal{O}_{L^1}\left(\frac{n}{\ln^{k+1}n}\right)$ nodes before the repayment of the loss of balls, these additional balls cannot give more than $\mathcal{O}_{L^1}\left(\frac{n}{\ln^k n}\right)$ nodes. Now we consider the vertices with $d \geq \mu^{-1} \ln n + \ln^{0.5+\epsilon} n$. Again we first distribute the n balls assuming that $s_1 = 0$, and then repay for the potential loss of balls in the subtrees if $s_1 > 0$. First note that for $d = \mathcal{O}(\ln n)$ we can argue as in the previous case. This means that the number of nodes with $\mu^{-1} \ln n + \ln^{0.5+\epsilon} n \leq d \leq K \ln n$ for some arbitrary constant K is bounded by $\mathcal{O}_{L^1}\left(\frac{n}{\ln^k n}\right)$. For larger d we argue as follows: For any constant $K_1 > 0$,

$$\begin{aligned} & \text{mBin}(s_1, \prod_{j=2}^d W_{j,v}) + \text{mBin}(s_1, \prod_{j=3}^d W_{j,v}) + \cdots + \text{mBin}(s_1, W_{d,v}) + s_1 \\ & \leq \text{mBin}(s_1, \prod_{j=2}^d W_{j,v}) + \cdots + \text{mBin}(s_1, \prod_{j=d-\lfloor K_1 \ln n \rfloor}^d W_{j,v}) + K_1 s_1 \ln n. \end{aligned}$$

The Markov inequality gives,

$$\begin{aligned} & \mathbf{P}(\text{mBin}(s_1, \prod_{j=2}^d W_{j,v}) + \cdots + \text{mBin}(s_1, \prod_{j=d-\lfloor K_1 \ln n \rfloor}^d W_{j,v}) \geq 1) \\ & \leq \mathbf{E}\left(\text{Bin}(s_1, \prod_{j=2}^d W_{j,v}) + \cdots + \text{Bin}(s_1, \prod_{j=d-\lfloor K_1 \ln n \rfloor}^d W_{j,v})\right) = \mathcal{O}\left(b^{-K_1 \ln n}\right), \end{aligned} \tag{85}$$

where the last equality is obtained by first condition on \mathcal{G}_d and then take the expected value twice. Thus, the expected number of vertices that gets a repayment of at least $K_1 s_1 \ln n + 2$ balls is bounded by $\mathcal{O}\left(\frac{n}{b^{K_1 \ln n}}\right)$. Since $s_1 > 0$, we can assume that $d \leq n$. Hence, the expected number of balls of

this contribution is $\mathcal{O}\left(\frac{n^2}{b^{K_1 \ln n}}\right)$; choosing K_1 large enough this number is just $o(1)$ and can thus be ignored.

It remains to prove that if $s_1 = 0$ the number of vertices v , where $d(v) \leq \mu^{-1} \ln n - \ln^{0.5+\epsilon} n$ or $d(v) \geq \mu^{-1} \ln n + \ln^{0.5+\epsilon} n$, with $n_v \geq 1$ is bounded by $\mathcal{O}_{L^1}\left(\frac{n}{\ln^{k+1} n}\right)$ for any constant k . Note that an upper bound of the expected number of vertices at depth d is given by

$$b^d \mathbf{P}(n_v \geq 2), \quad (86)$$

where v is a vertex at depth $d - 1$. Note that this is true even in the case $s_0 = 0$, since for all internal nodes $n_v \geq s + 1$. Choosing $t > 0$, an application of the Markov inequality implies that

$$\begin{aligned} \mathbf{P}(n_v \geq 2) &\leq \mathbf{P}(n_v(n_v - 1) \geq 2) \leq \\ &\mathbf{P}(n_v^t(n_v - 1)^t \geq 2^t) \leq \frac{\mathbf{E}(n_v^t(n_v - 1)^t)}{2^t}. \end{aligned} \quad (87)$$

Thus, an upper bound of the expected profile for the vertices at depth d is

$$b^d \mathbf{E}(n_v^t(n_v - 1)^t), \quad (88)$$

where v is a vertex at depth $d - 1$.

First we show that the number of vertices v (assuming $s_1 = 0$) where $d(v) \geq \mu^{-1} \ln n + \ln^{0.5+\epsilon} n$ is bounded by $\mathcal{O}_{L^1}\left(\frac{n}{\ln^{k+1} n}\right)$. We prove this by choosing $t = \frac{1+\epsilon(n)}{2}$, where $\epsilon(n) > 0$ is a decreasing function of n that we specify below, and show that

$$\sum_{d=\lfloor \mu^{-1} \ln n + \ln^{0.5+\epsilon} n \rfloor - 1}^{\infty} b^d \mathbf{E}\left(n_v^{\frac{1+\epsilon(n)}{2}}(n_v - 1)^{\frac{1+\epsilon(n)}{2}}\right) = \mathcal{O}\left(\frac{n}{\ln^{k+1} n}\right). \quad (89)$$

Let X_d be a mixed binomial $(n, \prod_{j=1}^d W_j)$, where W_j , $j \in 1, \dots, d$ are i.i.d. random variables distributed as V . To show (89) it is enough to show that the expected value of

$$\sum_{d=\lfloor \mu^{-1} \ln n + \ln^{0.5+\epsilon} n \rfloor - 1}^{\infty} b^d \mathbf{E}\left(X_d^{\frac{1+\epsilon(n)}{2}}(X_d - 1)^{\frac{1+\epsilon(n)}{2}} \middle| \mathcal{G}_d\right), \quad (90)$$

is $\mathcal{O}\left(\frac{n}{\ln^{k+1} n}\right)$. That this is enough follows because of the bound of n_v in (84), since we assume that $s_1 = 0$. Suppose that $\epsilon(n) < 1$, thus the Lyapounov

inequality (which is a special case of the well-known Hölder inequality) gives

$$\begin{aligned} \mathbf{E}\left(X_d^{\frac{1+\epsilon(n)}{2}}(X_d - 1)^{\frac{1+\epsilon(n)}{2}}|\mathcal{G}_d\right) &\leq (n^2 - n)^{\frac{1+\epsilon(n)}{2}} \prod_{j=1}^d W_j^{1+\epsilon(n)} \\ &\leq \left(n \prod_{j=1}^d W_j\right)^{1+\epsilon(n)}. \end{aligned} \quad (91)$$

Hence, to show (89) we deduce from the right hand-side of the second inequality in (91) that it is enough to show that

$$S_1 := \sum_{d=\lfloor \mu^{-1} \ln n + \ln^{0.5+\epsilon} n \rfloor - 1}^{\infty} b^d(\mathbf{E}(W_j^{1+\epsilon(n)}))^d n^{1+\epsilon(n)} = \mathcal{O}\left(\frac{n}{\ln^{k+1} n}\right). \quad (92)$$

Taylor expansion gives

$$\begin{aligned} W_j^{1+\epsilon(n)} &= W_j e^{\epsilon(n) \ln W_j} = W_j \left(1 + \epsilon(n) \ln W_j + \frac{\ln^2 W_j}{2} \epsilon^2(n) \right) \\ &\quad + \mathcal{O}\left(W_j \epsilon^3(n) \ln^3 W_j\right). \end{aligned} \quad (93)$$

Thus, by taking expectations in (93) we get

$$\begin{aligned} S_1 &= \sum_{d=\lfloor \mu^{-1} \ln n + \ln^{0.5+\epsilon} n \rfloor - 1}^{\infty} \left(1 - \mu \epsilon(n) + \frac{\sigma^2 + \mu^2}{2} \epsilon^2(n) + \mathcal{O}(\epsilon^3(n))\right)^d n^{1+\epsilon(n)} \\ &= \sum_{d=\lfloor \mu^{-1} \ln n + \ln^{0.5+\epsilon} n \rfloor - 1}^{\infty} e^{\ln \left(1 - \mu \epsilon(n) + \frac{\sigma^2 + \mu^2}{2} \epsilon^2(n) + \mathcal{O}(\epsilon^3(n))\right) d + \ln n (1 + \epsilon(n))} \\ &= \sum_{d=\lfloor \mu^{-1} \ln n + \ln^{0.5+\epsilon} n \rfloor - 1}^{\infty} e^{\left(-\mu \epsilon(n) + \frac{\sigma^2}{2} \epsilon^2(n) + \mathcal{O}(\epsilon^3(n))\right) d + \ln n (1 + \epsilon(n))} \\ &= \mathcal{O}\left(\frac{n^{1 - \mu \epsilon(n) \ln^{0.5+\epsilon} n + \mathcal{O}(\epsilon^2(n))}}{\epsilon(n)}\right). \end{aligned} \quad (94)$$

Hence, by choosing $\epsilon(n) = \delta \ln^{-0.5+\epsilon} n$ for some constant δ (that we choose small enough) we get from the last inequality in (94) that for some constant $B > 0$ and any constant k ,

$$S_1 = \mathcal{O}\left(n e^{-B \ln^{2\epsilon} n}\right) = \mathcal{O}\left(\frac{n}{\ln^{k+1} n}\right). \quad (95)$$

We argue similarly for the vertices v , $d(v) \leq \mu^{-1} \ln n - \ln^{0.5+\epsilon} n$. In (88) let $t = \frac{1-\epsilon(n)}{2}$ where $\epsilon(n) = \delta \ln^{-0.5+\epsilon} n$ for some constant δ as above. In analogy with (89) an upper bound for the expected number of vertices v with $d(v) \leq \mu^{-1} \ln n - \ln^{0.5+\epsilon} n$ is

$$S_2 := \sum_{d=0}^{\lfloor \mu^{-1} \ln n - \ln^{0.5+\epsilon} n \rfloor} b^d \mathbf{E}(n_v^{\frac{1-\epsilon(n)}{2}} (n_v - 1)^{\frac{1-\epsilon(n)}{2}}). \quad (96)$$

We use similar calculations as in (92)–(95) to show that

$$\sum_{d=0}^{\lfloor \mu^{-1} \ln n - \ln^{0.5+\epsilon} n \rfloor} b^d (\mathbf{E}(W_j^{1-\epsilon(n)}))^d n^{1-\epsilon(n)} = \mathcal{O}\left(n e^{-B \ln^{2\epsilon} n}\right). \quad (97)$$

This implies in analogy with (89)–(92) that for some constant B and any constant k ,

$$S_2 = \mathcal{O}(n e^{-B \ln^{2\epsilon} n}) = \mathcal{O}\left(\frac{n}{\ln^{k+1} n}\right). \quad (98)$$

Hence, if $s_1 = 0$ the number of bad vertices is bounded by $\mathcal{O}_{L^1}\left(\frac{n}{\ln^{k+1} n}\right)$, for any constant k . Thus, it follows from our previous explanation that the number of bad vertices for arbitrary $s_1 \geq 0$ is bounded by $\mathcal{O}_{L^1}\left(\frac{n}{\ln^k n}\right)$. \square

Remark 4.2. We note from (94), (95) and (98) that we in fact get a sharper bound for the number of bad nodes, i.e., $\mathcal{O}(n e^{-B' \ln^{2\epsilon} n})$ for some constant $B' > 0$.

Remark 4.3. From the calculations in the proof of Theorem 2.2 in particular in (94), we see that we can get a much smaller error term for larger depths, i.e., for any constant r there is a constant $C > 0$ so that the number of nodes with $d(v) \geq C \ln n$ is bounded by $\mathcal{O}_{L^1}\left(\frac{1}{n^r}\right)$.

4.3 Proof of Theorem 2.3

Proof. We write

$$Z_n := \frac{D_n - \mu^{-1} \ln n}{\sqrt{\ln n}}. \quad (99)$$

By a classical result in probability theory, see e.g. [10, Theorem 5.5.4], the limit law in (6) implies that (16) holds if Z_n is uniformly integrable. In particular this is true if Z_n^2 is uniformly integrable. This uniform integrability

also gives

$$\mathbf{E}(Z_n^2) := \frac{\mathbf{E}\left((D_n - \mu^{-1} \ln n)^2\right)}{\ln n} \rightarrow \mathbf{E}\left(N(0, \sigma^2 \mu^{-3})^2\right) = \sigma^2 \mu^{-3}. \quad (100)$$

Furthermore, the convergence results in (16) and (100) imply (18) for $k = n$. By using the same coupling argument as in (5) it is easy to show that the convergence result of the expected depth in (16) implies the convergence result of the expected average depth in (17).

Thus, it remains to show that Z_n^2 is uniformly integrable and that (18) for $k = n$ implies that (18) also holds for $\frac{n}{\ln n} \leq k < n$. By a standard argument, see e.g. [10, Theorem 5.5.4], Z_n^2 is uniformly integrable if for some $p > 1$ and n_0 large enough,

$$\sup_{n > n_0} \mathbf{E}(|Z_n^2|^p) := \sup_{n > n_0} \mathbf{E}\left(\left|\frac{(D_n - \mu^{-1} \ln n)^2}{\ln n}\right|^p\right), \quad (101)$$

is uniformly bounded. We choose $p = \frac{3}{2}$. We show that this is true by using similar calculations as Devroye used in [3] for proving the limit law of D_n in (6). First, consider an infinite random path u_1, u_2, \dots , in the skeleton tree S_b , where u_1 is the root. Given u_i and the split vector $\mathcal{V}_{u_i} = (V_1, \dots, V_b)$ for u_i , then u_{i+1} is the j -th child of i with probability V_j . Construct a random split tree with n balls and let u^* be the unique leaf in the infinite path. Then by using a natural coupling, letting the n -th ball follow the random path, D_n is in stochastic sense less than or equal to the distance between u^* and the root. In the coupling D_n is less than this distance, if the n -th ball is sent to a leaf which splits and does not send this ball to one of its children (i.e., the n -th ball is one of the s_0 balls). If the n -th ball is one of the s_1 balls it is added to a child of $p(u^*)$ (the parent of u^*), i.e., it ends up at the same depth as u^* . Recall that H_n denotes the height of a split tree with n balls. For all $\beta > 0$ we have

$$\mathbf{P}(D_n > k + \beta) \leq \mathbf{P}(n(u_k) > \beta) + \mathbf{P}(H_\beta > \beta), \quad (102)$$

and

$$\mathbf{P}(D_n < k) \leq \mathbf{P}(n(u_k) \leq s + 1). \quad (103)$$

Recall that $\Delta = V_S$, where given (V_1, \dots, V_b) , $S = j$ with probability V_j .

Then $n(u_k)$ is stochastically bounded by

$$\begin{aligned} n(u_k) &\leq \text{mBin}(n, \prod_{j=1}^k \Delta_j) + \text{mBin}(s_1, \prod_{j=2}^k \Delta_j) \\ &\quad + \text{mBin}(s_1, \prod_{j=3}^k \Delta_j) + \cdots + \text{mBin}(s_1, \Delta_k) + s_1, \end{aligned} \quad (104)$$

where Δ_j are i.i.d random variables distributed as Δ .

Consider the probability $\mathbf{P}(D_n > k + \beta)$, where $k = \lfloor \mu^{-1} \ln n + \frac{x}{2} \sqrt{\ln n} \rfloor$ for $x \in R^+$. We bound this by bounding the probabilities in the right hand-side of (102), choosing $\beta = \lfloor \frac{x}{2} \ln^{0.2}(n) \rfloor$. First note that the bound of $n(u_k)$ in (104) implies that in stochastic sense

$$\begin{aligned} n(u_k) &\leq \text{mBin}(n, \prod_{j=1}^k \Delta_j) + \text{mBin}(s_1, \prod_{j=1}^k \Delta_j) + \cdots \\ &\quad + \text{mBin}(s_1, \prod_{j=k-\lfloor \frac{x}{2} \ln^{0.1} n \rfloor + 1}^k \Delta_j) + \lfloor \frac{s_1 x}{2} \ln^{0.1} n \rfloor. \end{aligned} \quad (105)$$

Thus, we can bound the first probability in the right hand-side of (102) by

$$\begin{aligned} \mathbf{P}(n(u_k) > \beta) &\leq \mathbf{P}\left(\text{mBin}(n, \prod_{j=1}^k \Delta_j) + \lfloor \frac{s_1 x}{2} \ln^{0.1} n \rfloor \geq \beta - 1\right) \\ &\quad + \mathbf{P}\left(\text{mBin}(s_1, \prod_{j=1}^k \Delta_j) + \cdots + \text{mBin}(s_1, \prod_{j=k-\lfloor \frac{x}{2} \ln^{0.1} n \rfloor}^k \Delta_j) > 1\right). \end{aligned} \quad (106)$$

For bounding the first probability in the right hand-side of the inequality in (106), we use [3, Lemma 4] which states a general result for bounding tail probabilities for mixed binomial (m, Z) distributions where Z is a random variable, thus we obtain

$$\begin{aligned} P_1 &:= \mathbf{P}\left(\text{mBin}(n, \prod_{j=1}^k \Delta_j) > \beta - \lfloor \frac{s_1 x}{2} \ln^{0.1} n \rfloor - 1\right) \\ &\leq \mathbf{P}\left(\sum_{j=1}^k \ln \Delta_j > \ln\left(\frac{\beta - \lfloor \frac{s_1 x}{2} \ln^{0.1} n \rfloor - 1}{2n}\right)\right) + \left(\frac{e}{4}\right)^{\frac{\beta - \lfloor \frac{s_1 x}{2} \ln^{0.1} n \rfloor - 1}{2}}. \end{aligned} \quad (107)$$

From (107) we deduce that for n large enough

$$\begin{aligned} P_1 &\leq \mathbf{P}\left(\frac{\sum_{j=1}^k \ln \Delta_j + k\mu}{\sqrt{k\sigma^2}} > \frac{\ln\left(\frac{\beta - \lfloor \frac{s_1 x}{2} \ln^{0.1} n \rfloor - 1}{2n}\right) + k\mu}{\sqrt{k\sigma^2}}\right) + \left(\frac{e}{4}\right)^{\frac{\beta - \lfloor \frac{s_1 x}{2} \ln^{0.1} n \rfloor - 1}{2}} \\ &\leq \mathbf{P}\left(\frac{\sum_{j=1}^k \ln \Delta_j + k\mu}{\sqrt{k\sigma^2}} > \frac{x\mu^{\frac{3}{2}}}{3\sigma}\right) + \left(\frac{e}{4}\right)^{\frac{\beta - \lfloor \frac{s_1 x}{2} \ln^{0.1} n \rfloor - 1}{2}}. \end{aligned} \quad (108)$$

Recall the notations $c := \mathbf{E}(\Delta)$, $\mu =: \mathbf{E}(-\ln \Delta)$ and $\sigma^2 =: \mathbf{Var}(\ln \Delta)$. Note that $c < 1$. Since the Δ_j , $j \in \{1, \dots, k\}$, are i.i.d random variables we can use the Marcinkiewicz-Zygmund inequalities, see e.g. [10, Corollary 3.8.2], which gives for $q \geq 2$,

$$\mathbf{E}\left(\left|\sum_{j=1}^k \ln \Delta_j + k\mu\right|^q\right) \leq B_q k^{\frac{q}{2}} \mathbf{E}\left(\left|\ln \Delta_j + \mu\right|^q\right), \quad (109)$$

where B_q is a constant only depending on q . By using the Markov inequality and (109) we get from (108) that for n large enough

$$\begin{aligned} P_1 &\leq \frac{\mathbf{E}\left(\left(\frac{\sum_{j=1}^k \ln \Delta_j + k\mu}{\sqrt{k\sigma^2}}\right)^4\right)}{\left(\frac{x\mu^{\frac{3}{2}}}{3\sigma}\right)^4} + \left(\frac{e}{4}\right)^{\frac{\beta - \lfloor \frac{s_1 x}{2} \ln^{0.1} n \rfloor - 1}{2}} \\ &\leq \frac{B_4 \mathbf{E}\left(\left|\ln \Delta_j + \mu\right|^4\right)}{\left(\frac{x\mu^{\frac{3}{2}}}{3}\right)^4} + \left(\frac{e}{4}\right)^{\frac{\beta - \lfloor \frac{s_1 x}{2} \ln^{0.1} n \rfloor - 1}{2}} = \frac{C}{x^4} + \left(\frac{e}{4}\right)^{\frac{\beta - \lfloor \frac{s_1 x}{2} \ln^{0.1} n \rfloor - 1}{2}}, \end{aligned} \quad (110)$$

for the constant $C = \frac{B_4 \mathbf{E}(|\ln \Delta_j + \mu|^4)^{3^4}}{\mu^6} < \infty$ (recall from section 1.2 that all moments of $|\ln \Delta|$ are bounded). The Markov inequality implies that

$$\begin{aligned} &\mathbf{P}\left(\text{mBin}(s_1, \prod_{j=1}^k \Delta_j) + \dots + \text{mBin}(s_1, \prod_{j=k - \lfloor \frac{x}{2} \ln^{0.1} n \rfloor + 1}^k \Delta_j) \geq 1\right) \\ &\leq \mathbf{E}\left(\text{mBin}(s_1, \prod_{j=1}^k \Delta_j) + \dots + \text{mBin}(s_1, \prod_{j=k - \lfloor \frac{x}{2} \ln^{0.1} n \rfloor + 1}^k \Delta_j)\right) \\ &= \mathcal{O}\left(c^{\lfloor \frac{x}{2} \ln^{0.1} n \rfloor}\right), \quad \text{for } c := \mathbf{E}(\Delta) < 1. \end{aligned} \quad (111)$$

We now consider the other probability i.e., $\mathbf{P}(H_\beta > \beta)$. (Note that this probability is 0 if $s_0 > 0$ or $s_1 > 0$.) By applying (86) we get

$$\mathbf{P}(H_\beta > \beta) \leq b^\beta \mathbf{P}(n(v) \geq 2), \quad (112)$$

where v is a vertex at depth $\beta - 1$. From (87) we deduce for $t = 0.75$,

$$\mathbf{P}(n_v \geq 2) \leq \mathbf{E}(n_v^{0.75}(n_v - 1)^{0.75}). \quad (113)$$

Let X_β be a mixed binomial $(n, \prod_{j=1}^\beta W_j)$, where W_j , $j \in \{1, \dots, \beta\}$ are i.i.d. random variables distributed as V . Note similarly as in (90) that (113) is bounded by the expectation of

$$\mathbf{E}(X_\beta^{0.75}(X_\beta - 1)^{0.75} | \mathcal{G}_\beta). \quad (114)$$

We note similarly as in (91) that the Lyapounov inequality gives

$$\mathbf{E}(X_\beta^{0.75}(X_\beta - 1)^{0.75} | \mathcal{G}_\beta) \leq (\beta^2 - \beta)^{0.75} \prod_{j=1}^\beta W_j^{1.5} \leq \left(\beta \prod_{j=1}^\beta W_j\right)^{1.5}. \quad (115)$$

Again the fact that $\mathbf{E}(W_j^2) < \mathbf{E}(W_j) = \frac{1}{b}$ (since $W_j \in (0, 1)$), gives that there is a $\delta > 0$ such that

$$\mathbf{P}(H_\beta > \beta) \leq b^{-\delta\beta} \beta^{1.5}. \quad (116)$$

We now consider the probability $\mathbf{P}(D_n < k)$, where $k = \lfloor \mu^{-1} \ln n - x\sqrt{\ln n} \rfloor$ for $x \in R^+$, and use the bound of the larger probability in (103). We have

$$\mathbf{P}(n(u_k) \leq s + 1) \leq \mathbf{P}\left(-ks + \text{Bin}\left(n, \prod_{j=1}^k \Delta_j\right) \leq s + 1\right). \quad (117)$$

Again by applying [3, Lemma 4] and using similar calculations as in (107)–(110), we get for n large enough

$$\begin{aligned} P_2 &\leq \mathbf{P}\left(\frac{\sum_{j=1}^k \ln \Delta_j + k\mu}{\sqrt{k\sigma^2}} < \frac{\ln\left(\frac{2(s(k+1)+1)}{n}\right) + k\mu}{\sqrt{k\sigma^2}}\right) + \left(\frac{2}{e}\right)^{s(k+1)+1} \\ &\leq \mathbf{P}\left(\frac{\sum_{j=1}^k \ln \Delta_j + k\mu}{\sqrt{k\sigma^2}} < \frac{x\mu^{\frac{3}{2}}}{3\sigma}\right) + \left(\frac{2}{e}\right)^{s(k+1)+1} \\ &\leq \frac{B_4 \mathbf{E}\left(\left|\ln \Delta_j + \mu\right|^4\right)}{\left(\frac{x\mu^{\frac{3}{2}}}{3}\right)^4} + \left(\frac{2}{e}\right)^{s(k+1)+1} = C \frac{1}{x^4} + \left(\frac{2}{e}\right)^{s(k+1)+1}, \end{aligned} \quad (118)$$

for the constant $C = \frac{B_4 \mathbf{E}(|\ln \Delta_j + \mu|^4) 3^4}{\mu^6} < \infty$. Now we can show that for n_0 large enough $\sup_{n > n_0} \mathbf{E}(|Z_n^2|^{\frac{3}{2}})$ in (101) is uniformly bounded: By the choice of k and β , we get from (110), (111), (116) and (118) that for n_0 large enough

$$\begin{aligned} \sup_{n > n_0} \mathbf{E}(|Z_n^2|^{\frac{3}{2}}) &:= \sup_{n > n_0} \mathbf{E}\left(\left|\frac{(D_n - \mu^{-1} \ln n)^2}{\ln n}\right|^{\frac{3}{2}}\right) \\ &= \sup_{n > n_0} \int_{x=0}^{\infty} 3x^2 \mathbf{P}\left(\left|\frac{D_n - \mu^{-1} \ln n}{\sqrt{\ln n}}\right| > x\right) dx \\ &\leq \sup_{n > n_0} \left\{ \int_{x=1}^{\infty} \left(\frac{6C}{x^2} + 3x^2 \left(\frac{e}{4}\right)^{\frac{\beta - \lfloor \frac{s_1 x}{2} \ln^{0.1} n \rfloor - 1}{2}} + 3x^2 \left(\frac{2}{e}\right)^{s(k+1)+1}\right. \right. \\ &\quad \left. \left. + \mathcal{O}\left(x^2 c^{\lfloor \frac{x}{2} \ln^{0.1} n \rfloor}\right) + 3x^2 b^{-\delta\beta} \beta^{1.5}\right) dx \right\} + 1 < \infty, \end{aligned} \quad (119)$$

and thus Z_n^2 is uniformly integrable so that (100) holds, which shows (18) for $k = n$.

From this result it is now easy to show as we explain below that (18) also holds for all k , $\frac{n}{\ln n} \leq k < n$. Recall that we denote the depth of ball k , when it is added to the tree by D_k^f . As we argued for proving (5), in stochastic sense for $k \leq n$,

$$D_k^f \leq D_k \leq D_n. \quad (120)$$

From (6) it follows that for all $\frac{n}{\ln n} \leq k \leq n$,

$$\frac{D_k^f - \mu^{-1} \ln n}{\sqrt{\sigma^2 \mu^{-3} \ln n}} \xrightarrow{d} N(0, 1).$$

By using this and (120), for $\frac{n}{\ln n} \leq k \leq n$,

$$\frac{D_k - \mu^{-1} \ln n}{\sqrt{\sigma^2 \mu^{-3} \ln n}} \xrightarrow{d} N(0, 1).$$

We need to show that for $\frac{n}{\ln n} \leq k \leq n$,

$$\frac{\mathbf{E}\left((D_k - \mu^{-1} \ln n)^2\right)}{\ln n} \rightarrow \mathbf{E}\left(N(0, \sigma^2 \mu^{-3})^2\right). \quad (121)$$

As for D_n this follows if for n_0 large enough,

$$\sup_{n > n_0} \mathbf{E}\left(\left|\frac{(D_k - \mu^{-1} \ln n)^2}{\ln n}\right|^{\frac{3}{2}}\right) < \infty. \quad (122)$$

We have for $k \leq n$,

$$\begin{aligned} \mathbf{P}\left(\frac{D_k^f - \mu^{-1} \ln n}{\sqrt{\ln n}} \geq x\right) &\leq \mathbf{P}\left(\frac{D_k - \mu^{-1} \ln n}{\sqrt{\ln n}} \geq x\right) \\ &\leq \mathbf{P}\left(\frac{D_n - \mu^{-1} \ln n}{\sqrt{\ln n}} \geq x\right), \end{aligned}$$

and

$$\begin{aligned} \mathbf{P}\left(\frac{D_n - \mu^{-1} \ln n}{\sqrt{\ln n}} < x\right) &\leq \mathbf{P}\left(\frac{D_k - \mu^{-1} \ln n}{\sqrt{\ln n}} < x\right) \\ &\leq \mathbf{P}\left(\frac{D_k^f - \mu^{-1} \ln n}{\sqrt{\ln n}} < x\right). \end{aligned}$$

Thus, (122) follows from the calculations in (119). This shows that (18) holds for all k , $\frac{n}{\ln n} \leq k < n$, follows from the fact that (18) holds for $k = n$. \square

We now prove the two corollaries of Theorem 2.3.

Proof of Corollary 2.1. We show (20) and from this it is obvious that (19) also holds since Theorem 2.2 implies that the bad vertices are few enough so that we could equally sum over all vertices.

First note that (121) gives that for the balls $\lfloor \frac{n}{\ln n} \rfloor \leq k \leq n$,

$$\mathbf{E}(D_k - \mu^{-1} \ln n)^2 = \mu^{-3} \sigma^2 \ln n + o(\ln n). \quad (123)$$

Recall that a vertex v in a split tree T is called good if

$$\mu^{-1} \ln n - \ln^{0.5+\epsilon} n \leq d(v) \leq \mu^{-1} \ln n + \ln^{0.5+\epsilon} n,$$

and that we write $\mathbf{V}^*(T^n)$ for the set of good vertices in T^n and $N^* = |\mathbf{V}^*(T^n)|$ for the number of good vertices. Note that (14) and Theorem 2.2 implies that

$$\mathbf{E}(N^*) = \alpha n + o(n). \quad (124)$$

We will now consider subtrees defined similarly as the $T_{r,B}$, $r \in R$, subtrees we used in the proof of Theorem 2.1. However, instead of using the product M_v^N for defining the stopping time in each branch we use the real subtree size n_v : Let U be the set of vertices such that $u \in U$, if and only if $n_u \leq \ln^{0.4} n$ and $n_{p(u)} > \ln^{0.4} n$ (where $p(u)$ is the parent of u), and consider all subtrees $T_u^{\ln^{0.4} n}$, $u \in U$, rooted at u .

It is an immediate consequence of Lemma 4.1 that the first and second moment of the height of a subtree with $\ln^{0.4} n$ balls is bounded by $\mathcal{O}(\ln^{0.4} n)$ and $\mathcal{O}(\ln^{0.8} n)$, respectively. However, there are much stronger bounds, e.g., [3] since split trees are of logarithmic order. Hence, since the subtrees are small by applying that $d(v) - \mu^{-1} \ln n = (d(u) - \mu^{-1} \ln n) + (d(v) - d(u))$ and summing over all good vertices we get

$$\begin{aligned} & \mathbf{E} \left(\sum_u \sum_{v \in \mathbf{V}^*(T_u^{\ln^{0.4} n})} (d(v) - \mu^{-1} \ln n)^2 \right) \\ &= \mathbf{E} \left(\sum_u \sum_{v \in \mathbf{V}^*(T_u^{\ln^{0.4} n})} (d(u) - \mu^{-1} \ln n)^2 \right) + o(n \ln n). \end{aligned} \quad (125)$$

In (125) we use the bound for the good vertices, but it is obvious from Theorem 2.2 that the bad vertices are few enough so that one could equally sum over all vertices. The number of subtrees that hold the balls $k < \lfloor \frac{n}{\ln n} \rfloor$ is trivially bounded by $\lfloor \frac{n}{\ln n} \rfloor$. Thus, the number of nodes in these subtrees is bounded by $\mathcal{O}_{L^1} \left(\frac{n}{\ln^{0.6} n} \right)$. Let $N_u^* = |\mathbf{V}^*(T_u^{\ln^{0.4} n})|$ be the number of good vertices in $T_u^{\ln^{0.4} n}$. Hence, by applying that the subtrees $T_u^{\ln^{0.4} n}$, $u \in U$, are small so that $(d(v) - \mu^{-1} \ln n)^2$ do not differ more than $\mathcal{O}(\ln^{0.8} n)$ for different vertices $v \in T_u^{\ln^{0.4} n}$, together with (124) we get

$$\begin{aligned} & \mathbf{E} \left(\sum_u \sum_{v \in \mathbf{V}^*(T_u^{\ln^{0.4} n})} (d(u) - \mu^{-1} \ln n)^2 |n_u| \right) \\ &= \sum_u \mathbf{E} \left(\sum_{v \in \mathbf{V}^*(T_u^{\ln^{0.4} n})} \sum_{k \in T_u} \frac{(D_k - \mu^{-1} \ln n)^2}{n_u} |n_u| \right) + o_{L^1}(n \ln n) \\ &= \sum_u \sum_{k \in T_u} \frac{\mathbf{E} \left((D_k - \mu^{-1} \ln n)^2 |n_u| \right)}{n_u} (\alpha n_u + o(n_u)) + o_{L^1}(n \ln n) \end{aligned} \quad (126)$$

Recall from (45) in Lemma 4.3 that

$$\mathbf{E}(N) = \mathbf{E} \left(\sum_{r \in R} N_r I\{|n_r - M_r^n| \leq B^{0.6}\} \right) + \mathcal{O} \left(\frac{n}{B^{0.1}} \right),$$

where M_r^n by definition is less than B . If we choose $B := \ln^{0.3} n$ this means that for the expected value in the right hand-side we can assume that $n_r \leq$

$\ln^{0.4} n$. Hence, the expected number of (good) vertices in T^n that are not in the subtrees $T_u^{\ln^{0.4} n}$, $u \in U$, is bounded by $\mathcal{O}\left(\frac{n}{B^{0.1}}\right)$ for $B = \ln^{0.3} n$. Hence, this bound implies that the expected value of the last equality in (126) is equal to

$$\begin{aligned} & \sum_{k \in T^n} \mathbf{E}\left((D_k - \mu^{-1} \ln n)^2\right)(\alpha + o(1)) + o(n \ln n) \\ &= \left(\mu^{-3} \sigma^2 \ln n + o(\ln n)\right)(\alpha n + o(n)) + o(n \ln n). \end{aligned}$$

□

Proof of Corollary 2.2. As in Corollary 2.1 we only show the result for the good vertices, i.e., (22). From the proof it is obvious that also (21) holds by applying Theorem 2.2, showing that the number of bad vertices is covered by the error term. We observe the obvious fact that the sum of those n_i , $i \in \{1, \dots, b^L\}$, which are less than $\frac{n}{b^{kL}}$ for large enough k , is bounded by

$$b^L \cdot \frac{n}{b^{kL}} = \mathcal{O}\left(\frac{n}{\ln^4 n}\right). \quad (127)$$

(Note that by choosing k large enough in (127), the power of the logarithm can be arbitrarily large.)

Recall that $\mathbf{V}^*(T_i)$ is the set of good vertices in T_i and that Ω_L is the σ -field generated by $\{n_v, d(v) \leq L\}$. Let

$$Z_i := \sum_{v \in \mathbf{V}^*(T_i)} \frac{(d_i(v) - \mu^{-1} \ln n_i)^2}{\mu^{-3} \ln^3 n_i}.$$

Thus, from (20) it follows that

$$\sum_{i=1}^{b^L} \mathbf{E}\left(Z_i \middle| \Omega_L\right) = \sum_{i=1}^{b^L} \frac{\sigma^2 \alpha n_i}{\ln^2 n_i} + \sum_{i=1}^{b^L} \frac{o(n_i)}{\ln^2 n_i}.$$

Let $k > 0$ be a fixed constant and assume that n_i is at least $\frac{n}{b^{kL}}$; by Taylor expansion we get

$$\frac{1}{\ln^2 n_i} = \frac{1}{\ln^2 n} + \mathcal{O}\left(\frac{\ln \ln n}{\ln^3 n}\right). \quad (128)$$

By applying (127) for k large enough to cover those n_i that are less than $\frac{n}{b^{kL}}$ in an error term $o\left(\frac{n}{\ln^2 n}\right)$, and using (128) we deduce

$$\sum_{i=1}^{b^L} \frac{o(n_i)}{\ln^2 n_i} = o\left(\sum_{i=1}^{b^L} \frac{n_i}{\ln^2 n_i}\right) + o\left(\frac{n}{\ln^2 n}\right) = o\left(\frac{n}{\ln^2 n}\right). \quad (129)$$

Hence, since we can assume that n_i is at least $\frac{n}{b^{kL}}$ for large enough k , by Taylor expansion

$$\sum_{i=1}^{b^L} \frac{\sigma^2 \alpha n_i}{\ln^2 n_i} = \frac{\sigma^2 \alpha n}{\ln^2 n} + \mathcal{O}\left(\sum_{i=1}^{b^L} \frac{n_i \ln \ln n}{\ln^3 n}\right) = \frac{\sigma^2 \alpha n}{\ln^2 n} + o\left(\frac{n}{\ln^2 n}\right). \quad (130)$$

Since only the good vertices are considered, and the random variables conditioned on Ω_L are independent for $i \in \{1, \dots, b^L\}$,

$$\mathbf{Var}\left(\sum_{i=1}^{b^L} Z_i \middle| \Omega_L\right) = \sum_{i=1}^{b^L} \mathbf{Var}\left(Z_i \middle| \Omega_L\right) \leq \mu^3 \sum_{i=1}^{b^L} \mathbf{Var}\left(\sum_{v \in \mathbf{V}^*(T_i)} \frac{1}{\ln^{2-2\epsilon} n_i} \middle| \Omega_L\right). \quad (131)$$

Thus, the well-known Minkowski's inequality and the fact that $\mathbf{E}(N^2) = \mathcal{O}(n^2)$ imply

$$\mathbf{Var}\left(\sum_{i=1}^{b^L} Z_i \middle| \Omega_L\right) = \mathcal{O}\left(\sum_{i=1}^{b^L} \frac{n_i^2}{\ln^{4-4\epsilon} n_i}\right). \quad (132)$$

Similarly as in (59) for β large enough,

$$\sum_{i=1}^{b^L} \mathbf{E}(n_i^2) = o\left(\frac{n^2}{\ln^4 n}\right), \quad (133)$$

Applying (133), Chebyshev's inequality gives (22). □

5 Results on the Total Path Lengths

We complete this study with some results and a conjecture of the “total path length” random variables. Recall from Section 1.2 the definitions of the two types of total path length $\Psi(T)$ and $\Upsilon(T)$, i.e., the sum of the depths of balls and the sum of the depths of nodes, respectively.

From (17) we have

$$\mathbf{E}\left(\Psi(T^n)\right) = \mu^{-1} n \ln n + nq(n), \quad (134)$$

where $q(n) = o(\ln^{0.5} n)$ is a function that depends on the type of split tree.

Similarly, by using (14) in Theorem 2.1 and the profile result in Theorem 2.2 including Remark 4.3 (which gives a smaller bound of the expected number of vertices with depths much bigger than the depths of the good vertices), we get

$$\mathbf{E}\left(\Upsilon(T^n)\right) = \mu^{-1}\alpha n \ln n + nr(n), \quad (135)$$

where α is the constant that occurs in (14) and $r(n) = o(\ln n)$ is a function that depends on the type of split tree.

(A2). Assume that the functions $q(n)$ in (134) converges to some constant ς .

In [17] there is an analogous assumption. Examples of split trees where it is shown that $q(n)$ converges to a constant are binary search trees (e.g. [8]), random m -ary search trees [15], quad trees [17] and the random median of a $(2k+1)$ -tree [18], tries and Patricia tries [2].

(A3). We assume that the result in (14) in Theorem 2.1 can be improved to

$$\mathbf{E}(N) = \alpha n + f(n),$$

where $f(n) = \mathcal{O}\left(\frac{n}{\ln^{1+\epsilon} n}\right)$.

Stronger second order terms of the size have previously been shown to hold e.g., for m -ary search trees [16], for these $f(n)$ in assumption (A3) is $o(\sqrt{n})$ when $m \leq 26$ and is $\mathcal{O}(n^{1-\epsilon})$ when $m \geq 27$. Further, as described in Section 1.3 tries are special cases of split trees which are not as random as other types of split trees. Flajolet and Vallée (personal communication) have recently shown that also for most tries (as long as $-\ln V$ is not too close to being lattice) assumption (A3) holds.

Theorem 5.1. Assume that (A1)–(A3) hold, then also $r(n)$ converges to some constant ζ .

Let

$$\Gamma_n := \alpha n q(n) - nr(n), \quad (136)$$

and note that

$$\alpha \mathbf{E}\left(\Psi(T^n)\right) - \mathbf{E}\left(\Upsilon(T^n)\right) = \Gamma_n. \quad (137)$$

For proving Theorem 5.1 we will show that $\frac{\Gamma_n}{n}$ converges to a constant. We write \sum_v^* for a sum where we sum over all vertices $v \in T^n$ except the root i.e., $v \neq \sigma$. First we recall that the total pathlength for the balls is equivalent to the sum of all subtree sizes (except for the the whole tree) for the balls i.e.,

$$\Psi(T^n) = \sum_v^* n_v, \quad (138)$$

where σ is the root of T^n . Similarly we recall that the total pathlength for the nodes is equivalent to the sum of all subtree sizes (except for the the whole tree) for the nodes i.e.,

$$\Upsilon(T^n) = \sum_v^* N_v, \quad (139)$$

where σ is the root of T^n . Hence, by assuming (A3) we get from (137) that

$$\begin{aligned} \Gamma_n &= \alpha \mathbf{E} \left(\sum_v^* n_v \right) - \mathbf{E} \left(\sum_v^* \left(\alpha n_v + \mathcal{O} \left(\frac{n_v}{\ln^{1+\epsilon} n_v} \right) \right) \right) \\ &= \mathbf{E} \left(\sum_v^* \mathcal{O} \left(\frac{n_v}{\log^{1+\epsilon} n_v} \right) \right). \end{aligned} \quad (140)$$

We will again consider the $T_{r,B}$, $r \in R$, subtrees from the proof of Theorem 2.1 in Section 3 (defined such that $M_r^n := n \prod_{j=1}^{d(r)} W_j < B$ and $M_{p(r)}^n := n \prod_{j=1}^{d(r)-1} W_j \geq B$). However, here we choose B differently, i.e., $B = \epsilon^{-20}$.

Lemma 5.1. *Assume that (A1)–(A3) hold, then*

$$\alpha \mathbf{E} \left(\Psi(T^n) \right) - \mathbf{E} \left(\Upsilon(T^n) \right) = \Gamma_n = \mathcal{O}(n). \quad (141)$$

Furthermore,

$$\Gamma_n = \sum_{r \in R} \mathbf{E}(\Gamma_{n_r}) + o(n), \quad (142)$$

where

$$\Gamma_{n_r} = \alpha \mathbf{E} \left(\Psi(T_{r,B}) \middle| n_r \right) - \mathbf{E} \left(\Upsilon(T_{r,B}) \middle| n_r \right). \quad (143)$$

Proof. Assuming (A3), we get from (140) that

$$\begin{aligned}
\Gamma_n &= \mathbf{E} \left(\sum_v^* \mathcal{O} \left(\frac{n_v}{\log^{1+\epsilon} n_v} \right) \right) \\
&= \sum_k \mathbf{E} \left(\sum_{v: 2^k \leq n_v < 2^{k+1}}^* \mathcal{O} \left(\frac{n_v}{\log^{1+\epsilon} n_v} \right) \right) \\
&= \sum_k \mathcal{O} \left(\frac{n}{2^k} \cdot \frac{2^k}{k^{1+\epsilon}} \right) = \mathcal{O}(n),
\end{aligned} \tag{144}$$

where we applied Corollary 3.1 in the last equality. In the same way, for the vertices which are not in the $T_{r,B}$, $r \in R$, subtrees (ignoring the root σ) we deduce that

$$\mathbf{E} \left(\sum_{\substack{v \notin T_{r,B}, \ r \in R, \\ v \neq \sigma}} \mathcal{O} \left(\frac{n_v}{\log^{1+\epsilon} n_v} \right) \right) = \sum_{k \geq \epsilon^{-20}} \mathcal{O} \left(\frac{n}{2^k} \cdot \frac{2^k}{k^{1+\epsilon}} \right) = o(n). \tag{145}$$

Hence, (142) follows from (145) and (144). \square

Proof of Theorem 5.1. We will use the same type of proof as the proof of (14) in Theorem 2.1. We start with two arbitrary values of the cardinality n and \hat{n} , where $\hat{n} \geq n$, and show that

$$\left| \frac{\Gamma_n}{n} - \frac{\Gamma_{\hat{n}}}{\hat{n}} \right| = \mathcal{O}(\epsilon) \quad \text{as } n \rightarrow \infty, \quad \forall \epsilon > 0. \tag{146}$$

Since (146) implies that $\frac{\Gamma_n}{n}$ is Cauchy it also converges to some constant as n tends to infinity; hence, we deduce Theorem 5.1. Recall from the proof of Theorem 2.1 that a main application for the proof is to use (39) in Lemma 4.1. Here we use an analogous applications of (144) in Lemma 5.1, i.e., $\Gamma_n = \mathcal{O}(n)$.

Recall that we prove Lemma 4.3 by showing

$$\mathbf{E} \left(\sum_r n_r I \{ |n_r - M_r^n| \geq B^{0.6} \} \right) = \mathcal{O} \left(\frac{n}{B^{0.1}} \right), \tag{147}$$

and then applying (39) in Lemma 4.1. In the same way by using (147) and (141) as well as (142) in Lemma 5.1 we get that

$$\Gamma_n = \mathbf{E} \left(\sum_{r \in R} \Gamma_{n_r} I \{ |n_r - M_r^n| \leq B^{0.6} \} \right) + o(n) + \mathcal{O} \left(\frac{n}{B^{0.1}} \right). \tag{148}$$

Recall that $R' \subseteq R$ is the set of vertices such that $r \in R'$ if

$$|n_r - M_r^n| \leq B^{0.6}, \quad (149)$$

and that $R'' \subseteq R'$ is the set of vertices such that $r \in R''$ if $r \in R'$ and

$$\epsilon B < M_r^n < B. \quad (150)$$

Lemma 4.3 shows that we only need to consider the vertices in $r \in R'$. Similarly as in (50) we get

$$\Gamma_n = \mathbf{E}\left(\sum_{r \in R''} \Gamma_{n_r}\right) + \mathcal{O}(\epsilon n) + \mathcal{O}\left(\frac{n}{B^{0.1}}\right). \quad (151)$$

Recall from the proof of Theorem 2.1 that we sub-divide the $T_{r,B}$, $r \in R$, subtrees into smaller classes, wherein the M_r^n , $r \in R$, in each class are close to each-other. As before we choose $\gamma = \epsilon^2$, and let $Z = \{B, B - \gamma B, B - 2\gamma B, \dots, \epsilon B\}$, where $\epsilon = \frac{1}{k}$ for some positive integer k . Recall that we write $R_z \subseteq R$, $z \in Z$, for the set of vertices $r \in R$, such that $M_r^n \in [z - \gamma B, z)$ and $M_{p(r)}^n \geq B$. Hence, (151) gives

$$\Gamma_n = \mathbf{E}\left(\sum_{z \in Z} \sum_{r \in R' \cap R_z} \Gamma_{n_r}\right) + \mathcal{O}(\epsilon n) + \mathcal{O}\left(\frac{n}{B^{0.1}}\right). \quad (152)$$

To approximate the expected value in (152) we apply a lemma that is similar to Lemma 4.2.

Lemma 5.2. *Adding K balls to a tree can at most have an influence on $\mathbf{E}(\Psi(T^n))$ and $\mathbf{E}(\Upsilon(T^n))$, respectively, by $\mathcal{O}(K \ln(n + K))$.*

Proof. Adding one ball to a tree with n balls the expected depth is $\mathcal{O}(\ln n)$ and the expected number of additional nodes is $\mathcal{O}(1)$. Note that $\mathcal{O}(1)$ nodes only have distances $\mathcal{O}(1)$ between each other.) Hence, when the K -th ball is added the expected depth is $\mathcal{O}(\ln(n + K))$. Since K balls give an expectation of $\mathcal{O}(K)$ nodes the result holds for both $\mathbf{E}(\Psi(T^n))$ and $\mathbf{E}(\Upsilon(T^n))$. \square

Let r_z be an arbitrarily chosen node in $R' \cap R_z$, where $z \in Z$. Similarly as in (52), by using (149) and Lemma 5.2, from (152) we get

$$\Gamma_n = \sum_{z \in Z} \mathbf{E}(|R' \cap R_z|) \left(\Gamma_{n_{r_z}} + \mathcal{O}(\gamma B \ln B) \right) + \mathcal{O}(\epsilon n) + \mathcal{O}\left(\frac{n}{B^{0.1}}\right). \quad (153)$$

Define b_x in a tree with cardinality $\lfloor x \rfloor$ as $b_x := \frac{\Gamma_{\lfloor x \rfloor}}{\lfloor x \rfloor}$, and note from Lemma 5.1 that $b_x = \mathcal{O}(1)$. Recall that $S = \{1, 1 - \gamma, 1 - 2\gamma, \dots, \epsilon\}$, where $\gamma = \epsilon^2$.

Recall from (53) that for each choice of γ and $\alpha \in S$, there is a σ_γ such that for a constant c_α (depending on α),

$$\left| \frac{\mathbf{E}(|R' \cap R_{\alpha B}|)}{\frac{n}{B}} - c_\alpha \right| \leq \gamma^2 + \mathcal{O}\left(\frac{1}{B^{0.1}}\right), \quad (154)$$

whenever $\frac{n}{B} \geq \frac{1}{\sigma_\gamma}$. By choosing $B = \epsilon^{-20}$ for n large enough $\frac{n}{B} \geq \frac{1}{\sigma_\gamma}$ so that (154) holds. Moreover, since $\sum_{\alpha \in S} c_\alpha = \mathcal{O}(1)$ we have that $\sum_{\alpha \in S} c_\alpha \frac{\mathcal{O}(B\gamma \ln B)}{B} = \mathcal{O}(\gamma \ln \epsilon)$. Recall that $\gamma = \epsilon^2$. Thus, for a constant c_α (depending on α) and $b_{\alpha B} = \mathcal{O}(1)$, (153) and (154) imply that

$$\begin{aligned} \Gamma_n &= n \sum_{\alpha \in S} c_\alpha \frac{1}{B} \left(b_{\alpha B} \alpha B + \mathcal{O}(B\gamma \ln B) \right) + n \sum_{\alpha \in S} \mathcal{O}(b_{\alpha B} \gamma^2) + \mathcal{O}(\epsilon n) = \\ &= n \sum_{\alpha \in S} \alpha b_{\alpha B} c_\alpha + \mathcal{O}(\epsilon n). \end{aligned} \quad (155)$$

In analogy, also for $\hat{n} \geq n$,

$$\Gamma_{\hat{n}} = \hat{n} \sum_{\alpha \in S} \alpha b_{\alpha B} c_\alpha + \mathcal{O}(\epsilon \hat{n}). \quad (156)$$

Thus, (146) follows, which shows Theorem 5.1. \square

Finally we present a theorem that is applied in [12].

Theorem 5.2. *Let $L = \lfloor \beta \log_b \ln n \rfloor$ for some large enough constant β . Assume that (A1)–(A3) hold, then*

$$\sum_{i=1}^{b^L} \frac{\Psi(T_i)}{\mu^{-2} \ln^2 n_i} = \sum_{i=1}^{b^L} \frac{n_i}{\mu^{-1} \ln n_i} + \frac{n\zeta}{\mu^{-2} \ln^2 n} + o_p\left(\frac{n}{\ln^2 n}\right), \quad (157)$$

and

$$\sum_{i=1}^{b^L} \frac{\Upsilon(T_i)}{\mu^{-2} \ln^2 n_i} = \sum_{i=1}^{b^L} \frac{\alpha n_i}{\mu^{-1} \ln n_i} + \frac{n\zeta}{\mu^{-2} \ln^2 n} + o_p\left(\frac{n}{\ln^2 n}\right). \quad (158)$$

Proof. We only show (158), since we can use exactly the same type of arguments for showing (157).

First, (135) gives

$$\mathbf{E}\left(\sum_{i=1}^{b^L} \frac{\Upsilon(T_i)}{\mu^{-2} \ln^2 n_i} \middle| \Omega_L\right) = \sum_{i=1}^{b^L} \frac{\alpha n_i}{\mu^{-1} \ln n_i} + \sum_{i=1}^{b^L} \frac{n_i r(n_i)}{\mu^{-2} \ln^2 n_i}. \quad (159)$$

Note that conditioned on Ω_L , the summands $\Upsilon(T_i)$, $i \in \{1, \dots, b^L\}$ are independent. By applying the Cauchy-Schwarz inequality, and using the facts that $\mathbf{E}(N^2) = \mathcal{O}(n^2)$ and that $\mathbf{E}(D_k^2) = \mathcal{O}(\ln^2 n)$ for all k , we deduce that

$$\begin{aligned} \mathbf{Var}\left(\sum_{i=1}^{b^L} \Upsilon(T_i) \middle| \Omega_L\right) &= \sum_{i=1}^{b^L} \mathbf{Var}\left(\Upsilon(T_i) \middle| \Omega_L\right) \\ &\leq \sum_{i=1}^{b^L} \mathbf{E}\left(\Upsilon(T_i)^2 \middle| \Omega_L\right) = \sum_{i=1}^{b^L} \mathcal{O}(n_i^2 \ln^2 n_i). \end{aligned} \quad (160)$$

Similarly as in (133), for any constant k (and choosing the constant β in L large enough) the following holds

$$\mathbf{E}\left(\sum_{i=1}^{b^L} n_i^2\right) = \frac{n^2}{\ln^k n}. \quad (161)$$

Thus, for a large enough constant β , by taking expectations in (160) we get

$$\mathbf{EVar}\left(\sum_{i=1}^{b^L} \frac{\Upsilon(T_i)}{\mu^{-2} \ln^2 n_i} \middle| \Omega_L\right) = o\left(\frac{n^2}{\ln^4 n}\right). \quad (162)$$

Using (159) and (160) and applying (162), the Chebyshev inequality results in that conditioning on Ω_L ,

$$\sum_{i=1}^{b^L} \frac{\Upsilon(T_i)}{\mu^{-2} \ln^2 n_i} = \sum_{i=1}^{b^L} \frac{\alpha n_i}{\mu^{-1} \ln n_i} + \sum_{i=1}^{b^L} \frac{n_i r(n_i)}{\mu^{-2} \ln^2 n_i} + o_p\left(\frac{n}{\ln^2 n}\right). \quad (163)$$

By applying Theorem 5.1, (128) and (129) we get

$$\begin{aligned} \sum_{i=1}^{b^L} \frac{n_i r(n_i)}{\mu^{-2} \ln^2 n_i} &= \sum_{i=1}^{b^L} \frac{\zeta n_i}{\mu^{-2} \ln^2 n_i} + \sum_{i=1}^{b^L} \frac{o(n_i)}{\ln^2 n_i} \\ &= \frac{\zeta n}{\ln^2 n} + o\left(\frac{n}{\ln^2 n}\right). \end{aligned} \quad (164)$$

Thus, (158) follows from (163) and (164). □

Acknowledgement:

Professor Svante Janson is gratefully acknowledged for invaluable support and advice. I also thank Dr Nicolas Broutin for helpful discussions.

References

- [1] S. Asmussen, *Applied Probability and Queues*. John Wiley Sons, Chichester, 1987.
- [2] J. Bourdon, Size and path length of Patricia tries: dynamical sources context. *Random Structures Algorithms* **19** (2001), no. 3-4, 289–315.
- [3] L. Devroye, Universal limit laws for depths in random trees. *SIAM J. Comput.* **28** (1998), no 2, 409–432.
- [4] L. Devroye, Applications of Stein’s method in the analysis of random binary search trees. *Stein’s Method and Applications*, 47–297 (ed. Chen, Barbour) Inst. for Math. Sci. Lect. Notes Ser. **5**, World Scientific Press, Singapore, 2005.
- [5] W. Feller, Fluctuation theory and recurrent events. *Trans. Amer. Math. Soc.* **67** (1949) 98–119, .
- [6] W. Feller, *An Introduction to Probability Theory and Its Applications. Vol. 1.* 3rd ed., Wiley, New York, 1968.
- [7] W. Feller, *An Introduction to Probability Theory and Its Applications. Vol. II.* 2nd ed., Wiley, New York, 1971.
- [8] J. A. Fill and S. Janson, Quicksort asymptotics. *J. Algorithms* **44** (2002), 4–28.
- [9] A. Gut, *Stopped Random Walks*. Springer Verlag, New York, Berlin, Heidelberg, 1988.
- [10] A. Gut, *Probability: A Graduate Course*, Springer, New York, 2005.
- [11] C. Holmgren, Random records and cuttings in binary search trees. Accepted in *Combinat. Probab. Comput.* (2009).
- [12] C. Holmgren, A weakly 1-stable limiting distribution for the number of random records and cuttings in split trees. *Submitted for publication*.
- [13] S. Janson, T. Łuczak and A. Rucinski, *Random Graphs.*, Wiley, New York, 2000.
- [14] M. Régnier and P. Jacquet, New results on the size of tries. *IEEE Trans. Inform. Theory* **35** (1989), no. 1, 203–205.

- [15] H. Mahmoud, On the average internal path length of m -ary search trees. *Acta Inform.* **23** (1986), 111–117.
- [16] H. Mahmoud and B. Pittel, Analysis of the space of search trees under the random insertion algorithm. *J. Algorithms* **10** (1989), no. 1, 52–75.
- [17] R. Neininger and L. Rüschendorf, On the internal pathlength of d -dimensional quad trees. *Random Struct. Alg.* **15** (1999), no. 1, 25–41.
- [18] U. Roesler, On the analysis of stochastic divide and conquer algorithms. Average-case analysis of algorithms (Princeton, NJ, 1998), *Algorithmica* **29** (2001), no. 1-2, 238–261.